LECTURE 5

Stochastic Processes

> We may regard the present state of the universe as the effect of its past and the cause of its future. An intellect which at a certain moment would know all forces that set nature in motion, and all positions of all items of which nature is composed, if this intellect were also vast enough to submit these data to analysis, it would embrace in a single formula the movements of the greatest bodies of the universe and those of the tiniest atom; for such an intellect nothing would be uncertain and the future just like the past would be present before its eyes.[1]

In many problems that involve modeling the behavior of some system, we lack sufficiently detailed information to determine how the system behaves, or the behavior of the system is so complicated that an exact description of it becomes irrelevant or impossible. In that case, a probabilistic model is often useful.

Probability and randomness have many different philosophical interpretations, but, whatever interpretation one adopts, there is a clear mathematical formulation of probability in terms of measure theory, due to Kolmogorov.

Probability is an enormous field with applications in many different areas. Here we simply aim to provide an introduction to some aspects that are useful in applied mathematics. We will do so in the context of stochastic processes of a continuous time variable, which may be thought of as a probabilistic analog of deterministic ODEs. We will focus on Brownian motion and stochastic differential equations, both because of their usefulness and the interest of the concepts they involve.

Before discussing Brownian motion in Section 3, we provide a brief review of some basic concepts from probability theory and stochastic processes.

1. Probability

> Mathematicians are like Frenchmen: whatever you say to them they translate into their own language and forthwith it is something entirely different.[2]

A *probability space* $(\Omega, \mathcal{F}, P)$ consists of: (a) a sample space Ω, whose points label all possible outcomes of a random trial; (b) a σ-algebra $\mathcal{F}$ of measurable subsets of Ω, whose elements are the events about which it is possible to obtain information; (c) a probability measure $P : \mathcal{F} \to [0, 1]$, where $0 \leq P(A) \leq 1$ is the probability that the event $A \in \mathcal{F}$ occurs. If $P(A) = 1$, we say that an event A

[1] Pierre Simon Laplace, in *A Philosophical Essay on Probabilities.*

[2] Johann Goethe. It has been suggested that Goethe should have said "Probabilists are like Frenchmen (or Frenchwomen)."

occurs *almost surely*. When the σ-algebra $\mathcal{F}$ and the probability measure P are understood from the context, we will refer to the probability space as Ω.

In this definition, we say that$\mathcal{F}$ is σ-algebra on Ω if it is is a collection of subsets of Ω such that $\emptyset$ and Ω belong to $\mathcal{F}$, the complement of a set in $\mathcal{F}$ belongs to $\mathcal{F}$, and a countable union or intersection of sets in $\mathcal{F}$ belongs to $\mathcal{F}$. A probability measure P on $\mathcal{F}$ is a function $P : \mathcal{F} \to [0,1]$ such that $P(\emptyset) = 0$, $P(\Omega) = 1$, and for any sequence $\{A_n\}$ of pairwise disjoint sets (meaning that $A_i \cap A_j = \emptyset$ for $i \neq j$) we have

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P\left(A_n\right).$$

Example 5.1. Let Ω be a set and $\mathcal{F}$ a σ-algebra on Ω. Suppose that

$$\{\omega_n \in \Omega : n \in \mathbb{N}\}$$

is a countable subset of Ω and $\{p_n\}$ is a sequence of numbers $0 \leq p_n \leq 1$ such that $p_1 + p_2 + p_3 + \cdots = 1$. Then we can define a probability measure $P : \mathcal{F} \to [0,1]$ by

$$P(A) = \sum_{\omega_n \in A} p_n.$$

If $\mathcal{E}$ is a collection of subsets of a set Ω, then the σ-algebra generated by $\mathcal{E}$, denoted $\sigma(\mathcal{E})$, is the smallest σ-algebra that contains $\mathcal{E}$.

Example 5.2. The open subsets of $\mathbb{R}$ generate a σ-algebra $\mathcal{B}$ called the Borel σ-algebra of $\mathbb{R}$. This algebra is also generated by the closed sets, or by the collection of intervals. The interval $[0,1]$ equipped with the σ-algebra $\mathcal{B}$ of its Borel subsets and Lebesgue measure, which assigns to an interval a measure equal to its length, forms a probability space. This space corresponds to the random trial of picking a uniformly distributed real number from $[0,1]$.

1.1. Random variables

A function $X : \Omega \to \mathbb{R}$ defined on a set Ω with a σ-algebra $\mathcal{F}$ is said to be $\mathcal{F}$-measurable, or simply measurable when $\mathcal{F}$ is understood, if $X^{-1}(A) \in \mathcal{F}$ for every Borel set $A \in \mathcal{B}$ in $\mathbb{R}$. A *random variable* on a probability space $(\Omega, \mathcal{F}, P)$ is a real-valued $\mathcal{F}$-measurable function $X : \Omega \to \mathbb{R}$. Intuitively, a random variable is a real-valued quantity that can be measured from the outcome of a random trial.

If $f : \mathbb{R} \to \mathbb{R}$ is a Borel measurable function, meaning that $f^{-1}(A) \in \mathcal{B}$ for every $A \in \mathcal{B}$, and X is a random variable, then $Y = f \circ X$, defined by $Y(\omega) = f\left(X(\omega)\right)$, is also a random variable.

We denote the expected value of a random variable X with respect to the probability measure P by $\mathbf{E}^P[X]$, or $\mathbf{E}[X]$ when the measure P is understood. The expected value is a real number which gives the mean value of the random variable X. Here, we assume that X is *integrable*, meaning that the expected value $\mathbf{E}[\,|X|\,] < \infty$ is finite. This is the case if large values of X occur with sufficiently low probability.

Example 5.3. If X is a random variable with mean $\mu = \mathbf{E}[X]$, the *variance* σ^2 of X is defined by

$$\sigma^2 = \mathbf{E}\left[(X-\mu)^2\right],$$

assuming it is finite. The standard deviation σ provides a measure of the departure of X from its mean μ. The *covariance* of two random variables X_1, X_2 with means

μ_1, μ_2, respectively, is defined by

$$\operatorname{cov}(X_1, X_2) = \mathbf{E}\left[(X_1 - \mu_1)(X_2 - \mu_2)\right].$$

We will also loosely refer to this quantity as a correlation function, although strictly speaking the correlation function of X_1, X_2 is equal to their covariance divided by their standard deviations.

The expectation is a linear functional on random variables, meaning that for integrable random variables X, Y and real numbers c we have

$$\mathbf{E}[X+Y] = \mathbf{E}[X] + \mathbf{E}[Y], \qquad \mathbf{E}[cX] = c\mathbf{E}[X].$$

The expectation of an integrable random variable X may be expressed as an integral with respect to the probability measure P as

$$\mathbf{E}[X] = \int_\Omega X(\omega)\, dP(\omega).$$

In particular, the probability of an event $A \in \mathcal{F}$ is given by

$$P(A) = \int_A dP(\omega) = \mathbf{E}[1_A]$$

where $1_A : \Omega \to \{0,1\}$ is the indicator function of A,

$$1_A(\omega) = \begin{cases} 1 & \text{if } \omega \in A, \\ 0 & \text{if } \omega \notin A. \end{cases}$$

We will say that two random variables are equal P-almost surely, or almost surely when P is understood, if they are equal on an event A such that $P(A) = 1$. Similarly, we say that a random variable $X : A \subset \Omega \to \mathbb{R}$ is defined almost surely if $P(A) = 1$. Functions of random variables that are equal almost surely have the same expectations, and we will usually regard such random variables as being equivalent.

Suppose that $\{X_\lambda : \lambda \in \Lambda\}$ is a collection of functions $X_\lambda : \Omega \to \mathbb{R}$. The σ-algebra generated by $\{X_\lambda : \lambda \in \Lambda\}$, denoted $\sigma(X_\lambda : \lambda \in \Lambda)$, is the smallest σ-algebra $\mathcal{G}$ such that X_λ is $\mathcal{G}$-measurable for every $\lambda \in \Lambda$. Equivalently, $\mathcal{G} = \sigma(\mathcal{E})$ where $\mathcal{E} = \left\{X_\lambda^{-1}(A) : \lambda \in \Lambda,\, A \in \mathcal{B}(\mathbb{R})\right\}$.

1.2. Absolutely continuous and singular measures

Suppose that $P, Q : \mathcal{F} \to [0,1]$ are two probability measures defined on the same σ-algebra $\mathcal{F}$ of a sample space Ω.

We say that Q is *absolutely continuous* with respect to P is there is an integrable random variable $f : \Omega \to \mathbb{R}$ such that for every $A \in \mathcal{F}$ we have

$$Q(A) = \int_A f(\omega) dP(\omega).$$

We will write this relation as

$$dQ = f dP,$$

and call f the density of Q with respect to P. It is defined P-almost surely. In that case, if $\mathbf{E}^P$ and $\mathbf{E}^Q$ denote the expectations with respect to P and Q, respectively, and X is a random variable which is integrable with respect to Q, then

$$\mathbf{E}^Q[X] = \int_\Omega X\, dQ = \int_\Omega fX\, dP = \mathbf{E}^P[fX].$$

We say that probability measures P and Q on $\mathcal{F}$ are *singular* if there is an event $A \in \mathcal{F}$ such that $P(A) = 1$ and $Q(A) = 0$ (or, equivalently, $P(A^c) = 0$ and $Q(A^c) = 1$). This means that events which occur with finite probability with respect to P almost surely do not occur with respect to Q, and visa-versa.

Example 5.4. Let P be the Lebesgue probability measure on $([0,1], \mathcal{B})$ described in Example 5.2. If $f : [0,1] \to [0,\infty)$ is a nonnegative, integrable function with

$$\int_0^1 f(\omega)\, d\omega = 1,$$

where $d\omega$ denotes integration with respect to Lebesgue measure, then we can define a measure Q on $([0,1], \mathcal{B})$ by

$$Q(A) = \int_A f(\omega)\, d\omega.$$

The measure Q is absolutely continuous with respect to P with density f. Note that P is not necessarily absolutely continuous with respect to Q; this is the case only if $f \neq 0$ almost surely and $1/f$ is integrable. If R is a measure on $([0,1], \mathcal{B})$ of the type given in Example 5.1 then R and P (or R and Q) are singular because the Lebesgue measure of any countable set is equal to zero.

1.3. Probability densities

The distribution function $F : \mathbb{R} \to [0,1]$ of a random variable $X : \Omega \to \mathbb{R}$ is defined by $F(x) = P\{\omega \in \Omega : X(\omega) \leq x\}$ or, in more concise notation,

$$F(x) = P\{X \leq x\}.$$

We say that a random variable is continuous if the probability measure it induces on $\mathbb{R}$ is absolutely continuous with respect to Lebesgue measure.[3] Most of the random variables we consider here will be continuous.

If X is a continuous random variable with distribution function F, then F is differentiable and

$$p(x) = F'(x)$$

is the probability density function of X. If $A \in \mathcal{B}(\mathbb{R})$ is a Borel subset of $\mathbb{R}$, then

$$P\{X \in A\} = \int_A p(x)\, dx.$$

The density satisfies $p(x) \geq 0$ and

$$\int_{-\infty}^{\infty} p(x)\, dx = 1.$$

Moreover, if $f : \mathbb{R} \to \mathbb{R}$ is any Borel-measurable function such that $f(X)$ is integrable, then

$$\mathbf{E}[f(X)] = \int_{-\infty}^{\infty} f(x)p(x)\, dx.$$

Example 5.5. A random variable X is Gaussian with mean μ and variance σ^2 if it has the probability density

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-(x-\mu)^2/\left(2\sigma^2\right)}.$$

[3]This excludes, for example, counting-type random variables that take only integer values.

We say that random variables $X_1, X_2, \ldots X_n : \Omega \to \mathbb{R}$ are jointly continuous if there is a joint probability density function $p(x_1, x_2, \ldots, x_n)$ such that

$$P\{X_1 \in A_1,\, X_1 \in A_1, \ldots,\, X_n \in A_n\} = \int_A p(x_1, x_2 \ldots, x_n)\; dx_1 dx_2 \ldots dx_n.$$

where $A = A_1 \times A_2 \times \cdots \times A_n$. Then $p(x_1, x_2, \ldots, x_n) \geq 0$ and

$$\int_{\mathbb{R}^n} p(x_1, x_2, \ldots, x_n)\; dx_1 dx_2 \ldots dx_n = 1.$$

Expected values of functions of the X_i are given by

$$\mathbf{E}[f(X_1, X_2, \ldots, X_n)] = \int_{\mathbb{R}^n} f(x_1, x_2, \ldots, x_n)\, p(x_1, x_2, \ldots, x_n)\; dx_1 dx_2 \ldots dx_n.$$

We can obtain the joint probability density of a subset of the X_i's by integrating out the other variables. For example, if $p(x, y)$ is the joint probability density of random variables X and Y, then the marginal probability densities $p_X(x)$ and $p_Y(y)$ of X and Y, respectively, are given by

$$p_X(x) = \int_{-\infty}^{\infty} p(x, y)\, dy, \qquad p_Y(y) = \int_{-\infty}^{\infty} p(x, y)\, dx.$$

Of course, in general, we cannot obtain the joint density $p(x, y)$ from the marginal densities $p_X(x)$, $p_Y(y)$, since the marginal densities do not contain any information about how X and Y are related.

Example 5.6. A random vector $\vec{X} = (X_1, \ldots, X_n)$ is Gaussian with mean $\vec{\mu} = (\mu_1, \ldots, \mu_n)$ and invertible covariance matrix $C = (C_{ij})$, where

$$\mu_i = \mathbf{E}[X_i], \qquad C_{ij} = \mathbf{E}[(X_i - \mu_i)(X_j - \mu_j)],$$

if it has the probability density

$$p(\vec{x}) = \frac{1}{(2\pi)^{n/2} (\det C)^{1/2}} \exp\left\{ -\frac{1}{2} (\vec{x} - \vec{\mu})^\top C^{-1} (\vec{x} - \vec{\mu}) \right\}.$$

Gaussian random variables are completely specified by their mean and covariance.

1.4. Independence

Random variables $X_1, X_2, \ldots, X_n : \Omega \to \mathbb{R}$ are said to be *independent* if

$$P\{X_1 \in A_1,\, X_2 \in A_2,\, \ldots,\, X_n \in A_n\} = P\{X_1 \in A_1\}\, P\{X_2 \in A_2\} \ldots P\{X_n \in A_n\}$$

for arbitrary Borel sets $A_1,\, A_2, \ldots, A_3 \subset \mathbb{R}$. If $X_1,\, X_2, \ldots,\, X_n$ are independent random variables, then

$$\mathbf{E}[f_1(X_1)\, f_2(X_2) \ldots f_n(X_n)] = \mathbf{E}[f_1(X_1)]\, \mathbf{E}[f_2(X_2)] \ldots \mathbf{E}[f_n(X_n)].$$

Jointly continuous random variables are independent if their joint probability density distribution factorizes into a product:

$$p(x_1, x_2, \ldots, x_n) = p_1(x_1)\, p_2(x_2) \ldots p_n(x_n).$$

If the densities $p_i = p_j$ are the same for every $1 \leq i, j \leq n$, then we say that $X_1, X_2, \ldots, X_n$ are independent, identically distributed random variables.

Heuristically, each random variable in a collection of independent random variables defines a different 'coordinate axis' of the probability space on which they are defined. Thus, any probability space that is rich enough to support a countably infinite collection of independent random variables is necessarily 'infinite-dimensional.'

Example 5.7. The Gaussian random variables in Example 5.6 are independent if and only if the covariance matrix C is diagonal.

The sum of independent Gaussian random variables is a Gaussian random variable whose mean and variance are the sums of those of the independent Gaussians. This is most easily seen by looking at the characteristic function of the sum,

$$\mathbf{E}\left[e^{i\xi(X_1+\cdots+X_n)}\right] = \mathbf{E}\left[e^{i\xi X_1}\right]\ldots\mathbf{E}\left[e^{i\xi X_n}\right],$$

which is the Fourier transform of the density. The characteristic function of a Gaussian with mean μ and variance σ^2 is $e^{i\xi\mu-\sigma^2\xi^2/2}$, so the means and variances add when the characteristic functions are multiplied. Also, a linear transformations of Gaussian random variables is Gaussian.

1.5. Conditional expectation

Conditional expectation is a somewhat subtle topic. We give only a brief discussion here. See [**45**] for more information and proofs of the results we state here.

First, suppose that $X : \Omega \to \mathbb{R}$ is an integrable random variable on a probability space $(\Omega, \mathcal{F}, P)$. Let $\mathcal{G} \subset \mathcal{F}$ be a σ-algebra contained in $\mathcal{F}$. Then the conditional expectation of X given $\mathcal{G}$ is a $\mathcal{G}$-measurable random variable

$$\mathbf{E}\left[X \mid \mathcal{G}\right] : \Omega \to \mathbb{R}$$

such that for all bounded $\mathcal{G}$-measurable random variables Z

$$\mathbf{E}\left[\,\mathbf{E}\left[\,X \mid \mathcal{G}\right] Z\,\right] = \mathbf{E}\left[XZ\right].$$

In particular, choosing $Z = 1_B$ as the indicator function of $B \in \mathcal{G}$, we get

$$\int_B \mathbf{E}\left[X \mid \mathcal{G}\right] dP = \int_B X\, dP \qquad \text{for all } B \in \mathcal{G}. \tag{5.1}$$

The existence of $\mathbf{E}\left[X \mid \mathcal{G}\right]$ follows from the Radon-Nikodym theorem or by a projection argument. The conditional expectation is only defined up to almost-sure equivalence, since (5.1) continues to hold if $\mathbf{E}[X \mid \mathcal{G}]$ is modified on an event in $\mathcal{G}$ that has probability zero. Any equations that involve conditional expectations are therefore understood to hold almost surely.

Equation (5.1) states, roughly, that $\mathbf{E}\left[X \mid \mathcal{G}\right]$ is obtained by averaging X over the finer σ-algebra $\mathcal{F}$ to get a function that is measurable with respect to the coarser σ-algebra $\mathcal{G}$. Thus, one may think of $\mathbf{E}\left[X \mid \mathcal{G}\right]$ as providing the 'best' estimate of X given information about the events in $\mathcal{G}$.

It follows from the definition that if X, XY are integrable and Y is $\mathcal{G}$-measurable then

$$\mathbf{E}\left[XY \mid \mathcal{G}\right] = Y\mathbf{E}\left[X \mid \mathcal{G}\right].$$

Example 5.8. The conditional expectation given the full σ-algebra $\mathcal{F}$, corresponding to complete information about events, is $\mathbf{E}\left[X \mid \mathcal{F}\right] = X$. The conditional expectation given the trivial σ-algebra $\mathcal{M} = \{\emptyset, \Omega\}$, corresponding to no information about events, is the constant function $\mathbf{E}\left[X \mid \mathcal{G}\right] = \mathbf{E}[X]$.

Example 5.9. Suppose that $\mathcal{G} = \{\emptyset, B, B^c, \Omega\}$ where B is an event such that $0 < P(B) < 1$. This σ-algebra corresponds to having information about whether or not the event B has occurred. Then

$$\mathbf{E}\left[X \mid \mathcal{G}\right] = p1_B + q1_{B^c}$$

where p, q are the expected values of X on B, B^c, respectively

$$p = \frac{1}{P(B)} \int_B X \, dP, \qquad q = \frac{1}{P(B^c)} \int_{B^c} X \, dP.$$

Thus, $\mathbf{E}[X \mid \mathcal{G}](\omega)$ is equal to the expected value of X given B if $\omega \in B$, and the expected value of X given B^c if $\omega \in B^c$.

The conditional expectation has the following 'tower' property regarding the collapse of double expectations into a single expectation: If $\mathcal{H} \subset \mathcal{G}$ are σ-algebras, then

$$\mathbf{E}[\mathbf{E}[X \mid \mathcal{G}] \mid \mathcal{H}] = \mathbf{E}[\mathbf{E}[X \mid \mathcal{H}] \mid \mathcal{G}] = \mathbf{E}[X \mid \mathcal{H}], \tag{5.2}$$

sometimes expressed as 'the coarser algebra wins.'

If $X, Y : \Omega \to \mathbb{R}$ are integrable random variables, we define the conditional expectation of X given Y by

$$\mathbf{E}[X \mid Y] = \mathbf{E}[X \mid \sigma(Y)].$$

This random variable depends only on the events that Y defines, not on the values of Y themselves.

Example 5.10. Suppose that $Y : \Omega \to \mathbb{R}$ is a random variable that attains countably many distinct values y_n. The sets $B_n = Y^{-1}(y_n)$, form a countable disjoint partition of Ω. For any integrable random variable X, we have

$$\mathbf{E}[X \mid Y] = \sum_{n \in \mathbb{N}} z_n \, 1_{B_n}$$

where 1_{B_n} is the indicator function of B_n, and

$$z_n = \frac{\mathbf{E}[1_{B_n} X]}{P(B_n)} = \frac{1}{P(B_n)} \int_{B_n} X \, dP$$

is the expected value of X on B_n. Here, we assume that $P(B_n) \neq 0$ for every $n \in \mathbb{N}$. If $P(B_n) = 0$ for some n, then we omit that term from the sum, which amounts to defining $\mathbf{E}[X \mid Y](\omega) = 0$ for $\omega \in B_n$. The choice of a value other than 0 for $\mathbf{E}[X \mid Y]$ on B_n would give an equivalent version of the conditional expectation. Thus, if $Y(\omega) = y_n$ then $\mathbf{E}[X \mid Y](\omega) = z_n$ where z_n is the expected value of $X(\omega')$ over all ω' such that $Y(\omega') = y_n$. This expression for the conditional expectation does not apply to continuous random variables Y, since then $P\{Y = y\} = 0$ for every $y \in \mathbb{R}$, but we will give analogous results below for continuous random variables in terms of their probability densities.

If $Y, Z : \Omega \to \mathbb{R}$ are random variables such that Z is measurable with respect to $\sigma(Y)$, then one can show that there is a Borel function $\varphi : \mathbb{R} \to \mathbb{R}$ such that $Z = \varphi(Y)$. Thus, there is a Borel function $\varphi : \mathbb{R} \to \mathbb{R}$ such that

$$\mathbf{E}[X \mid Y] = \varphi(Y).$$

We then define the conditional expectation of X given that $Y = y$ by

$$\mathbf{E}[X \mid Y = y] = \varphi(y).$$

Since the conditional expectation $\mathbf{E}[X \mid Y]$ is, in general, defined almost surely, we cannot define $\mathbf{E}[X \mid Y = y]$ unambiguously for all $y \in \mathbb{R}$, only for $y \in A$ where A is a Borel subset of $\mathbb{R}$ such that $P\{Y \in A\} = 1$.

More generally, if $Y_1, \ldots, Y_n$ are random variables, we define the conditional expectation of an integrable random variable X given $Y_1, \ldots, Y_n$ by

$$\mathbf{E}\left[X \mid Y_1, \ldots, Y_n\right] = \mathbf{E}\left[X \mid \sigma\left(Y_1, \ldots, Y_n\right)\right].$$

This is a random variable $\mathbf{E}\left[X \mid Y_1, \ldots, Y_n\right] : \Omega \to \mathbb{R}$ which is measurable with respect to $\sigma\left(Y_1, \ldots, Y_n\right)$ and defined almost surely. As before, there is a Borel function $\varphi : \mathbb{R}^n \to \mathbb{R}$ such that $\mathbf{E}\left[X \mid Y_1, \ldots, Y_n\right] = \varphi\left(Y_1, \ldots, Y_n\right)$. We denote the corresponding conditional expectation of X given that $Y_1 = y_1, \ldots, Y_n = y_n$ by

$$\mathbf{E}\left[X \mid Y_1 = y_1, \ldots, Y_n = y_n\right] = \varphi\left(y_1, \ldots, y_n\right).$$

Next we specialize these results to the case of continuous random variables. Suppose that $X_1, \ldots, X_m, Y_1, \ldots, Y_n$ are random variables with a joint probability density $p\left(x_1, x_2, \ldots x_m, y_1, y_2, \ldots, y_n\right)$. The conditional joint probability density of X_1, X_2,..., X_m given that $Y_1 = y_1$, $Y_2 = y_2$,..., $Y_n = y_n$, is

$$p\left(x_1, x_2, \ldots x_m \mid y_1, y_2, \ldots, y_n\right) = \frac{p\left(x_1, x_2, \ldots x_m, y_1, y_2, \ldots, y_n\right)}{p_Y\left(y_1, y_2, \ldots, y_n\right)}, \tag{5.3}$$

where p_Y is the marginal density of the $(Y_1, \ldots, Y_n)$,

$$p_Y\left(y_1, \ldots, y_n\right) = \int_{\mathbb{R}^m} p\left(x_1, \ldots, x_m, y_1, \ldots, y_n\right)\, dx_1 \ldots dx_m.$$

The conditional expectation of $f\left(X_1, \ldots, X_m\right)$ given that $Y_1 = y_1, \ldots, Y_n = y_n$ is

$$\begin{aligned}&\mathbf{E}\left[f\left(X_1, \ldots, X_m\right) \mid Y_1 = y_1, \ldots, Y_n = y_n\right]\\ &\quad = \int_{\mathbb{R}^m} f\left(x_1, \ldots, x_m\right) p\left(x_1, \ldots x_m \mid y_1, \ldots, y_n\right)\, dx_1, \ldots, dx_m.\end{aligned}$$

The conditional probability density $p\left(x_1, \ldots x_m \mid y_1, \ldots, y_n\right)$ in (5.3) is defined for $(y_1, \ldots, y_n) \in A$, where $A = \{(y_1, \ldots, y_n) \in \mathbb{R}^n : p_Y\left(y_1, \ldots, y_n\right) > 0\}$. Since

$$P\left\{(Y_1, \ldots, Y_n) \in A^c\right\} = \int_{A^c} p_Y(y_1, \ldots, y_n)\, dy_1 \ldots dy_n = 0$$

we have $P\{(Y_1, \ldots, Y_n) \in A\} = 1$.

Example 5.11. If X, Y are random variables with joint probability density $p(x, y)$, then the conditional probability density of X given that $Y = y$, is defined by

$$p(x \mid y) = \frac{p(x, y)}{p_Y(y)}, \qquad p_Y(y) = \int_{-\infty}^{\infty} p(x, y)\, dx,$$

provided that $p_Y(y) > 0$. Also,

$$\mathbf{E}\left[f(X, Y) \mid Y = y\right] = \int_{-\infty}^{\infty} f(x, y) p(x \mid y)\, dx = \frac{\int_{-\infty}^{\infty} f(x, y) p(x, y)\, dx}{p_Y(y)}.$$

2. Stochastic processes

Consider a real-valued quantity that varies 'randomly' in time. For example, it could be the brightness of a twinkling star, a velocity component of the wind at a weather station, a position or velocity coordinate of a pollen grain in Brownian motion, the number of clicks recorded by a Geiger counter up to a given time, or the value of the Dow-Jones index.

We describe such a quantity by a measurable function

$$X : [0, \infty) \times \Omega \to \mathbb{R}$$

where Ω is a probability space, and call X a stochastic process. The quantity $X(t,\omega)$ is the value of the process at time t for the outcome $\omega \in \Omega$. When it is not necessary to refer explicitly to the dependence of $X(t,\omega)$ on ω, we will write the process as $X(t)$. We consider processes that are defined on $0 \le t < \infty$ for definiteness, but one can also consider processes defined on other time intervals, such as $[0,1]$ or $\mathbb{R}$. One can also consider discrete-time processes with $t \in \mathbb{N}$, or $t \in \mathbb{Z}$, for example. We will consider only continuous-time processes.

We may think of a stochastic process in two different ways. First, fixing $\omega \in \Omega$, we get a function of time

$$X^\omega : t \mapsto X(t,\omega),$$

called a sample function (or sample path, or realization) of the process. From this perspective, the process is a collection of functions of time $\{X^\omega : \omega \in \Omega\}$, and the probability measure is a measure on the space of sample functions.

Alternatively, fixing $t \in [0,\infty)$, we get a random variable

$$X_t : \omega \mapsto X(t,\omega)$$

defined on the probability space Ω. From this perspective, the process is a collection of random variables $\{X_t : 0 \le t < \infty\}$ indexed by the time variable t. The probability measure describes the joint distribution of these random variables.

2.1. Distribution functions

A basic piece of information about a stochastic process X is the probability distribution of the random variables X_t for each $t \in [0,\infty)$. For example if X_t is continuous, we can describe its distribution by a probability density $p(x,t)$. These one-point distributions do not, however, tell us how the values of the process at different times are related.

Example 5.12. Let X be a process such that with probability $1/2$, we have $X_t = 1$ for all t, and with probability $1/2$, we have $X_t = -1$ for all t. Let Y be a process such that Y_t and Y_s are independent random variables for $t \neq s$, and for each t, we have $Y_t = 1$ with probability $1/2$ and $Y_t = -1$ with probability $1/2$. Then X_t, Y_t have the same distribution for each $t \in \mathbb{R}$, but they are different processes, because the values of X at different times are completely correlated, while the values of Y are independent. As a result, the sample paths of X are constant functions, while the sample paths of Y are almost surely discontinuous at every point (and non-Lebesgue measurable). The means of these processes, $\mathbf{E}X_t = \mathbf{E}Y_t = 0$, are equal and constant, but they have different covariances

$$\mathbf{E}\left[X_s X_t\right] = 1, \qquad \mathbf{E}\left[Y_s Y_t\right] = \begin{cases} 1 & \text{if } t = s, \\ 0 & \text{otherwise.} \end{cases}$$

To describe the relationship between the values of a process at different times, we need to introduce multi-dimensional distribution functions. We will assume that the random variables associated with the process are continuous.

Let $0 \le t_1 < t_2 < \cdots < t_n$ be a sequence times, and A_1, A_2,... A_n a sequence of Borel subsets $\mathbb{R}$. Let E be the event

$$E = \left\{\omega \in \Omega : X_{t_j}(\omega) \in A_j \text{ for } 1 \le j \le n\right\}.$$

Then, assuming the existence of a joint probability density $p\,(x_n, t_;\ldots;x_2, t_2; x_1, t_1)$ for X_{t_1}, X_{t_2},..., X_{t_n}, we can write

$$P\{E\} = \int_A p\,(x_n, t_n;\ldots;x_2, t_2; x_1, t_1)\; dx_1 dx_2 \ldots dx_n$$

where $A = A_1 \times A_2 \times \cdots \times A_n \subset \mathbb{R}^n$. We adopt the convention that times are written in increasing order from right to left in p.

These finite-dimensional densities must satisfy a consistency condition relating the $(n+1)$-dimensional densities to the n-dimensional densities: If $n \in \mathbb{N}$, $1 \leq i \leq n$ and $t_1 < t_2 < \cdots < t_i < \cdots < t_n$, then

$$\int_{-\infty}^{\infty} p\,(x_{n+1}, t_{n+1};\ldots;x_{i+1}, t_{i+1}; x_i, t_i; x_{i-1}, t_{i-1};\ldots;x_1, t_1)\; dx_i$$
$$= p\,(x_{n+1}, t_{n+1};\ldots;x_{i+1}, t_{i+1}; x_{i-1}, t_{i-1};\ldots;x_1, t_1)\,.$$

We will regard these finite-dimensional probability densities as providing a full description of the process. For continuous-time processes this requires an assumption of separability, meaning that the process is determined by its values at countably many times. This is the case, for example, if its sample paths are continuous, so that they are determined by their values at all rational times.

Example 5.13. To illustrate the inadequacy of finite-dimensional distributions for the description of non-separable processes, consider the process $X : [0,1] \times \Omega \to \mathbb{R}$ defined by

$$X(t,\omega) = \begin{cases} 1 & \text{if } t = \omega, \\ 0 & \text{otherwise,} \end{cases}$$

where $\Omega = [0,1]$ and P is Lebesgue measure on Ω. In other words, we pick a point $\omega \in [0,1]$ at random with respect to a uniform distribution, and change X_t from zero to one at $t = \omega$. The single time distribution of X_t is given by

$$P\,\{X_t \in A\} = \begin{cases} 1 & \text{if } 0 \in A, \\ 0 & \text{otherwise,} \end{cases}$$

since the probability that $\omega = t$ is zero. Similarly,

$$P\,\{X_{t_1} \in A_1, \ldots, X_{t_n} \in A_n\} = \begin{cases} 1 & \text{if } 0 \in \bigcap_{i=1}^n A_i, \\ 0 & \text{otherwise,} \end{cases}$$

since the probability that $\omega = t_i$ for some $1 \leq i \leq n$ is also zero. Thus, X has the same finite-dimensional distributions as the trivial zero-process $Z(t,\omega) = 0$. If, however, we ask for the probability that the realizations are continuous, we get different answers:

$$P\,\{X^\omega \text{ is continuous on } [0,1]\} = 0, \qquad P\,\{Z^\omega \text{ is continuous on } [0,1]\} = 1.$$

The problem here is that in order to detect the discontinuity in a realization X^ω of X, one needs to look at its values at an uncountably infinite number of times. Since measures are only countably additive, we cannot determine the probability of such an event from the probability of events that depend on the values of X^ω at a finite or countably infinite number of times.

2.2. Stationary processes

A process X_t, defined on $-\infty < t < \infty$, is *stationary* if X_{t+c} has the same distribution as X_t for all $-\infty < c < \infty$; equivalently this means that all of its finite-dimensional distributions depend only on time differences. 'Stationary' here is used in a probabilistic sense; it does not, of course, imply that the individual sample functions do not vary in time. For example, if one considers the fluctuations of a thermodynamic quantity, such as the pressure exerted by a gas on the walls of its container, this quantity varies in time even when the system is in thermodynamic equilibrium. The one-point probability distribution of the quantity is independent of time, but the two-point correlation at different times depends on the time difference.

2.3. Gaussian processes

A process is *Gaussian* if all of its finite-dimensional distributions are multivariate Gaussian distributions. A separable Gaussian process is completely determined by the means and covariance matrices of its finite-dimensional distributions.

2.4. Filtrations

Suppose that $X : [0, \infty) \times \Omega \to \mathbb{R}$ is a stochastic process on a probability space Ω with σ-algebra $\mathcal{F}$. For each $0 \leq t < \infty$, we define a σ-algebra $\mathcal{F}_t$ by

$$\mathcal{F}_t = \sigma\left(X_s : 0 \leq s \leq t\right). \tag{5.4}$$

If $0 \leq s < t$, then $\mathcal{F}_s \subset \mathcal{F}_t \subset \mathcal{F}$. Such a family of σ-fields $\{\mathcal{F}_t : 0 \leq t < \infty\}$ is called a *filtration* of $\mathcal{F}$.

Intuitively, $\mathcal{F}_t$ is the collection of events whose occurrence can be determined from observations of the process up to time t, and an $\mathcal{F}_t$-measurable random variable is one whose value can be determined by time t. If X is any random variable, then $\mathbf{E}\left[\, X \mid \mathcal{F}_t \,\right]$ is the 'best' estimate of X based on observations of the process up to time t.

The properties of conditional expectations with respect to filtrations define various types of stochastic processes, the most important of which for us will be Markov processes.

2.5. Markov processes

A stochastic process X is said to be a *Markov process* if for any $0 \leq s < t$ and any Borel measurable function $f : \mathbb{R} \to \mathbb{R}$ such that $f(X_t)$ has finite expectation, we have

$$\mathbf{E}\left[f\left(X_t\right) \mid \mathcal{F}_s\right] = \mathbf{E}\left[f\left(X_t\right) \mid X_s\right].$$

Here $\mathcal{F}_s$ is defined as in (5.4). This property means, roughly, that 'the future is independent of the past given the present.' In anthropomorphic terms, a Markov process only cares about its present state, and has no memory of how it got there.

We may also define a Markov process in terms of its finite-dimensional distributions. As before, we consider only processes for which the random variables X_t are continuous, meaning that their distributions can be described by probability densities. For any times

$$0 \leq t_1 < t_2 < \cdots < t_m < t_{m+1} < \cdots < t_n,$$

the conditional probability density that $X_{t_i} = x_i$ for $m+1 \le i \le n$ given that $X_{t_i} = x_i$ for $1 \le i \le m$ is given by

$$p\left(x_n, t_n; \ldots; x_{m+1}, t_{m+1} \mid x_m, t_m; \ldots; x_1, t_1\right) = \frac{p\left(x_n, t_n; \ldots; x_1, t_1\right)}{p\left(x_m, t_m; \ldots; x_1, t_1\right)}.$$

The process is a Markov process if these conditional densities depend only on the conditioning at the most recent time, meaning that

$$p\left(x_{n+1}, t_{n+1} \mid x_n, t_n; \ldots; x_2, t_2; x_1, t_1\right) = p\left(x_{n+1}, t_{n+1} \mid x_n, t_n\right).$$

It follows that, for a Markov process,

$$p\left(x_n, t_n; \ldots; x_2, t_2 \mid x_1, t_1\right) = p\left(x_n, t_n \mid x_{n-1}, t_{n-1}\right) \ldots p\left(x_2, t_2 \mid x_1, t_1\right).$$

Thus, we can determine all joint finite-dimensional probability densities of a continuous Markov process X_t in terms of the transition density $p\left(x, t \mid y, s\right)$ and the probability density $p_0(x)$ of its initial value X_0. For example, the one-point density of X_t is given by

$$p(x,t) = \int_{-\infty}^{\infty} p\left(x, t \mid y, 0\right) p_0(y)\, dy.$$

The transition probabilities of a Markov process are not arbitrary and satisfy the *Chapman-Kolmogorov equation*. In the case of a continuous Markov process, this equation is

$$p(x,t \mid y,s) = \int_{-\infty}^{\infty} p(x,t \mid z,r) p(z,r \mid y,s)\, dz \qquad \text{for any } s < r < t, \tag{5.5}$$

meaning that in going from y at time s to x at time t, the process must go though some point z at any intermediate time r.

A continuous Markov process is *time-homogeneous* if

$$p(x,t \mid y,s) = p(x,t-s \mid y,0),$$

meaning that its stochastic properties are invariant under translations in time. For example, a stochastic differential equation whose coefficients do not depend explicitly on time defines a time-homogeneous continuous Markov process. In that case, we write $p(x,t \mid y,s) = p(x,t-s \mid y)$ and the Chapman-Kolmogorov equation (5.5) becomes

$$p(x,t \mid y) = \int_{-\infty}^{\infty} p(x,t-s \mid z) p(z,s \mid y)\, dz \qquad \text{for any } 0 < s < t. \tag{5.6}$$

Nearly all of the processes we consider will be time-homogeneous.

2.6. Martingales

Martingales are fundamental to the analysis of stochastic processes, and they have important connections with Brownian motion and stochastic differential equations. Although we will not make use of them, we give their definition here.

We restrict our attention to processes M with continuous sample paths on a probability space $(\Omega, \mathcal{F}_t, P)$, where $\mathcal{F}_t = \sigma\left(M_t : t \ge 0\right)$ is the filtration induced by M. Then M is a *martingale*[4] if M_t has finite expectation for every $t \ge 0$ and for

[4]The term 'martingale' was apparently used in 18th century France as a name for the roulette betting 'strategy' of doubling the bet after every loss. If one were to compile a list of nondescriptive and off-putting names for mathematical concepts, 'martingale' would almost surely be near the top.

any $0 \leq s < t$,

$$\mathbf{E}[\, M_t \mid \mathcal{F}_s \,] = M_s.$$

Intuitively, a martingale describes a 'fair game' in which the expected value of a player's future winnings M_t is equal to the player's current winnings M_s. For more about martingales, see [**46**], for example.

3. Brownian motion

> The grains of pollen were particles...of a figure between cylindrical and oblong, perhaps slightly flattened...While examining the form of these particles immersed in water, I observed many of them very evidently in motion; their motion consisting not only of a change in place in the fluid manifested by alterations in their relative positions...In a few instances the particle was seen to turn on its longer axis. These motions were such as to satisfy me, after frequently repeated observations, that they arose neither from currents in the fluid, nor from its gradual evaporation, but belonged to the particle itself.[5]

In 1827, Robert Brown observed that tiny pollen grains in a fluid undergo a continuous, irregular movement that never stops. Although Brown was perhaps not the first person to notice this phenomenon, he was the first to study it carefully, and it is now known as Brownian motion.

The constant irregular movement was explained by Einstein (1905) and the Polish physicist Smoluchowski (1906) as the result of fluctuations caused by the bombardment of the pollen grains by liquid molecules. (It is not clear that Einstein was initially aware of Brown's observations — his motivation was to look for phenomena that could provide evidence of the atomic nature of matter.)

For example, a colloidal particle of radius $10^{-6}\,\mathrm{m}$ in a liquid, is subject to approximately 10^{20} molecular collisions each second, each of which changes its velocity by an amount on the order of $10^{-8}\,\mathrm{m\,s^{-1}}$. The effect of such a change is imperceptible, but the cumulative effect of an enormous number of impacts leads to observable fluctuations in the position and velocity of the particle.[6]

Einstein and Smoluchowski adopted different approaches to modeling this problem, although their conclusions were similar. Einstein used a general, probabilistic argument to derive a diffusion equation for the number density of Brownian particles as a function of position and time, while Smoluchowski employed a detailed kinetic model for the collision of spheres, representing the molecules and the Brownian particles. These approaches were partially connected by Langevin (1908) who introduced the Langevin equation, described in Section 5 below.

Perrin (1908) carried out experimental observations of Brownian motion and used the results, together with Einstein's theoretical predictions, to estimate Avogadro's number N_A; he found $N_A \approx 7 \times 10^{23}$ (see Section 6.2). Thus, Brownian motion provides an almost direct observation of the atomic nature of matter.

Independently, Louis Bachelier (1900), in his doctoral dissertation, introduced Brownian motion as a model for asset prices in the French bond market. This work received little attention at the time, but there has been extensive subsequent use of

[5]Robert Brown, from *Miscellaneous Botanical Works* Vol. I, 1866.

[6]Deutsch (1992) suggested that these fluctuations are in fact too small for Brown to have observed them with contemporary microscopes, and that the motion Brown saw had some other cause.

the theory of stochastic processes to model financial markets, especially following the development of the Black-Scholes-Merton (1973) model for options pricing (see Section 8).

Wiener (1923) gave the first construction of Brownian motion as a measure on the space of continuous functions, now called Wiener measure. Wiener did this by several different methods, including the use of Fourier series with random coefficients (*c.f.* (5.7) below). This work was further developed by Wiener and many others, especially Lévy (1939).

3.1. Definition

Standard (one-dimensional) *Brownian motion* starting at 0, also called the *Wiener process*, is a stochastic process $B(t, \omega)$ with the following properties:

(1) $B(0, \omega) = 0$ for every $\omega \in \Omega$;
(2) for every $0 \leq t_1 < t_2 < t_3 < \cdots < t_n$, the increments
$$B_{t_2} - B_{t_1}, \quad B_{t_3} - B_{t_2}, \ldots, \quad B_{t_n} - B_{t_{n-1}}$$
are independent random variables;
(3) for each $0 \leq s < t < \infty$, the increment $B_t - B_s$ is a Gaussian random variable with mean 0 and variance $t - s$;
(4) the sample paths $B^\omega : [0, \infty) \to \mathbb{R}$ are continuous functions for every $\omega \in \Omega$.

The existence of Brownian motion is a non-trivial fact. The main issue is to show that the Gaussian probability distributions, which imply that $B(t+\Delta t) - B(t)$ is typically of the order $\sqrt{\Delta t}$, are consistent with the continuity of sample paths. We will not give a proof here, or derive the properties of Brownian motion, but we will describe some results which give an idea of how it behaves. For more information on the rich mathematical theory of Brownian motion, see for example [**15, 46**].

The Gaussian assumption must, in fact, be satisfied by any process with independent increments and continuous sample sample paths. This is a consequence of the central limit theorem, because each increment
$$B_t - B_s = \sum_{i=0}^{n} \left(B_{t_{i+1}} - B_{t_i} \right) \qquad s = t_0 < t_1 < \cdots < t_n = t,$$
is a sum of arbitrarily many independent random variables with zero mean; the continuity of sample paths is sufficient to ensure that the hypotheses of the central limit theorem are satisfied. Moreover, since the means and variances of independent Gaussian variables are additive, they must be linear functions of the time difference. After normalization, we may assume that the mean of $B_t - B_s$ is zero and the variance is $t - s$, as in standard Brownian motion.

Remark 5.14. A probability distribution F is said to be *infinitely divisible* if, for every $n \in \mathbb{N}$, there exists a probability distribution F_n such that if $X_1, \ldots X_n$ are independent, identically distributed random variables with distribution F_n, then $X_1 + \cdots + X_n$ has distribution F. The Gaussian distribution is infinitely divisible, since a Gaussian random variable with mean μ and variance σ^2 is a sum of n independent, identically distributed random variables with mean μ/n and variance σ^2/n, but it is not the only such distribution; the Poisson distribution is another basic example. One can construct a stochastic process with independent increments for any infinitely divisible probability distribution. These processes are called Lévy

processes [**5**]. Brownian motion is, however, the only Lévy process whose sample paths are almost surely continuous; the paths of other Lévy processes contain jump discontinuities in any time interval with nonzero probability.

Since Brownian motion is a sum of arbitrarily many independent increments in any time-interval, it has a random fractal structure in which any part of the motion, after rescaling, has the same distribution as the original motion (see Figure 1). Specifically, if $c > 0$ is a constant, then

$$\tilde{B}_t = \frac{1}{c^{1/2}} B_{ct}$$

has the same distribution as B_t, so it is also a Brownian motion. Moreover, we may translate a Brownian motion B_t from any time s back to the origin to get a Brownian motion $\hat{B}_t = B_{t+s} - B_s$, and then rescale the translated process.

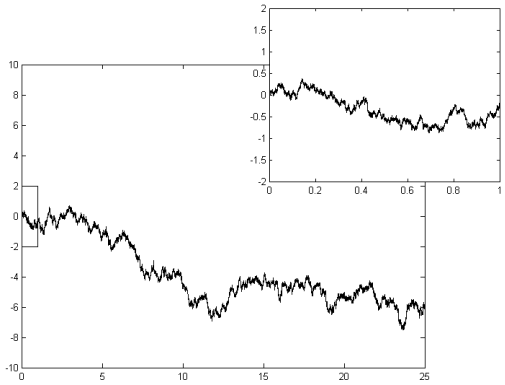

FIGURE 1. A sample path for Brownian motion, and a rescaling of it near the origin to illustrate the random fractal nature of the paths.

The condition of independent increments implies that Brownian motion is a Gaussian Markov process. It is not, however, stationary; for example, the variance t of B_t is not constant and grows linearly in time. We will discuss a closely related process in Section 5, called the stationary Ornstein-Uhlenbeck process, which is a stationary, Gaussian, Markov process (in fact, it is the only such process in one space dimension with continuous sample paths).

One way to think about Brownian motion is as a limit of random walks in discrete time. This provides an analytical construction of Brownian motion, and can be used to simulate it numerically. For example, consider a particle on a line that starts at $x = 0$ when $t = 0$ and moves as follows: After each time interval of length Δt, it steps a random distance sampled from independent, identically distributed Gaussian distributions with mean zero and variance Δt. Then, according to Donsker's theorem, the random walk approaches a Brownian motion in distribution as $\Delta t \to 0$. A key point is that although the total distance moved by the particle after time t goes to infinity as $\Delta t \to 0$, since it takes roughly on the order of $1/\Delta t$ steps of size $\sqrt{\Delta t}$, the net distance traveled remains finite almost surely

because of the cancelation between forward and backward steps, which have mean zero.

Another way to think about Brownian motion is in terms of random Fourier series. For example, Wiener (1923) showed that if $A_0, A_1, \ldots, A_n, \ldots$ are independent, identically distributed Gaussian variables with mean zero and variance one, then the Fourier series

$$B(t) = \frac{1}{\sqrt{\pi}} \left(A_0 t + 2 \sum_{n=1}^{\infty} A_n \frac{\sin nt}{n} \right) \tag{5.7}$$

almost surely has a subsequence of partial sums that converges uniformly to a continuous function. Furthermore, the resulting process B is a Brownian motion on $[0, \pi]$. The n^{th} Fourier coefficient in (5.7) is typically of the order $1/n$, so the uniform convergence of the series depends essentially on the cancelation between terms that results from the independence of their random coefficients.

3.2. Probability densities and the diffusion equation

Next, we consider the description of Brownian motion in terms of its finite-dimensional probability densities. Brownian motion is a time-homogeneous Markov process, with transition density

$$p(x, t \mid y) = \frac{1}{\sqrt{2\pi t}} \, e^{-(x-y)^2/2t} \qquad \text{for } t > 0. \tag{5.8}$$

As a function of (x, t), the transition density satisfies the diffusion, or heat, equation

$$\frac{\partial p}{\partial t} = \frac{1}{2} \frac{\partial^2 p}{\partial x^2}, \tag{5.9}$$

and the initial condition

$$p(x, 0 \mid y) = \delta(x - y).$$

The one-point probability density for Brownian motion starting at 0 is the Green's function of the diffusion equation,

$$p(x, t) = \frac{1}{\sqrt{2\pi t}} \, e^{-x^2/2t}.$$

More generally, if a Brownian motion B_t does not start almost surely at 0 and the initial value B_0 is a continuous random variable, independent of the rest of the motion, with density $p_0(x)$, then the density of B_t for $t > 0$ is given by

$$p(x, t) = \frac{1}{\sqrt{2\pi t}} \int e^{-(x-y)^2/2t} p_0(y) \, dy. \tag{5.10}$$

This is the Green's function representation of the solution of the diffusion equation (5.9) with initial data $p(x, 0) = p_0(x)$.

One may verify explicitly that the transition density (5.8) satisfies the Chapman-Kolmogorov equation (5.6). If we introduce the solution operators of (5.9),

$$T_t : p_0(\cdot) \mapsto p(\cdot, t)$$

defined by (5.10), then the Chapman-Kolmogorov equation is equivalent to the semi-group property $T_t T_s = T_{t+s}$. We use the term 'semi-group' here, because we cannot, in general, solve the diffusion equation backward in time, so T_t does not have an inverse (as would be required in a group).

The covariance function of Brownian motion is given by

$$\mathbf{E}\left[B_t B_s\right] = \min(t, s). \tag{5.11}$$

To see this, suppose that $s < t$. Then the increment $B_t - B_s$ has zero mean and is independent of B_s, and B_s has variance s, so

$$\mathbf{E}\left[B_s B_t\right] = \mathbf{E}\left[\left(B_t - B_s\right) B_s\right] + \mathbf{E}\left[B_s^2\right] = s.$$

Equivalently, we may write (5.11) as

$$\mathbf{E}\left[B_s B_t\right] = \frac{1}{2}\left(|t| + |s| - |t-s|\right).$$

Remark 5.15. One can a define a Gaussian process X_t, depending on a parameter $0 < H < 1$, called *fractional Brownian motion* which has mean zero and covariance function

$$\mathbf{E}\left[X_s X_t\right] = \frac{1}{2}\left(|t|^{2H} + |s|^{2H} - |t-s|^{2H}\right).$$

The parameter H is called the Hurst index of the process. When $H = 1/2$, we get Brownian motion. This process has similar fractal properties to standard Brownian motion because of the scaling-invariance of its covariance [**17**].

3.3. Sample path properties

Although the sample paths of Brownian motion are continuous, they are almost surely non-differentiable at every point.

We can describe the non-differentiablity of Brownian paths more precisely. A function $F : [a,b] \to \mathbb{R}$ is Hölder continuous on the interval $[a,b]$ with exponent γ, where $0 < \gamma \leq 1$, if there exists a constant C such that

$$|F(t) - F(s)| \leq C|t-s|^{\gamma} \qquad \text{for all } s,t \in [a,b].$$

For $0 < \gamma < 1/2$, the sample functions of Brownian motion are almost surely Hölder continuous with exponent γ on every bounded interval; but for $1/2 \leq \gamma \leq 1$, they are almost surely not Hölder continuous with exponent γ on any bounded interval.

One way to understand these results is through the law of the iterated logarithm, which states that, almost surely,

$$\limsup_{t\to 0^+} \frac{B_t}{\left(2t\log\log\frac{1}{t}\right)^{1/2}} = 1, \qquad \liminf_{t\to 0^+} \frac{B_t}{\left(2t\log\log\frac{1}{t}\right)^{1/2}} = -1.$$

Thus, although the typical fluctuations of Brownian motion over times Δt are of the order $\sqrt{\Delta t}$, there are rare deviations which are larger by a very slowly growing, but unbounded, double-logarithmic factor of $\sqrt{2\log\log(1/\Delta t)}$.

Although the sample paths of Brownian motion are almost surely not Hölder continuous with exponent $1/2$, there is a sense in which Brownian motion satisfies a stronger condition probabilistically: When measured with respect to a given, non-random set of partitions, the quadratic variation of a Brownian path on an interval of length t is almost surely equal to t. This property is of particular significance in connection with Itô's theory of stochastic differential equations (SDEs).

In more detail, suppose that $[a,b]$ is any time interval, and let $\{\Pi_n : n \in \mathbb{N}\}$ be a sequence of non-random partitions of $[a,b]$,

$$\Pi_n = \{t_0, t_1, \ldots, t_n\}, \qquad a = t_0 < t_1 < \cdots < t_n = b.$$

To be specific, suppose that Π_n is obtained by dividing $[a,b]$ into n subintervals of equal length (the result is independent of the choice of the partitions, provided they are not allowed to depend on $\omega \in \Omega$ so they cannot be 'tailored' to fit each

realization individually). We define the *quadratic variation* of a sample function B_t on the time-interval $[a, b]$ by

$$QV_a^b(B_t) = \lim_{n\to\infty} \sum_{i=1}^{n} \left(B_{t_i} - B_{t_{i-1}}\right)^2.$$

The n terms in this sum are independent, identically distributed random variables with mean $(b-a)/n$ and variance $2(b-a)^2/n^2$. Thus, the sum has mean $(b-a)$ and variance proportional to $1/n$. Therefore, by the law of large numbers, the limit exists almost surely and is equal to $(b-a)$. By contrast, the quadratic variation of any continuously differentiable function, or any function of bounded variation, is equal to zero.

This property of Brownian motion leads to the formal rule of the Itô calculus that

$$(dB)^2 = dt. \tag{5.12}$$

The apparent peculiarity of this formula, that the 'square of an infinitesimal' is another first-order infinitesimal, is a result of the nonzero quadratic variation of the Brownian paths.

The Hölder continuity of the Brownian sample functions for $0 < \gamma < 1/2$ implies that, for any $\alpha > 2$, the α-variation is almost surely equal to zero:

$$\lim_{n\to\infty} \sum_{i=1}^{n} \left|B_{t_i} - B_{t_{i-1}}\right|^{\alpha} = 0.$$

3.4. Wiener measure

Brownian motion defines a probability measure on the space $C[0,\infty)$ of continuous functions, called Wiener measure, which we denote by W.

A cylinder set $\mathcal{C}$ is a subset of $C[0,\infty)$ of the form

$$\mathcal{C} = \left\{B \in C[0,\infty) : B_{t_j} \in A_j \text{ for } 1 \le j \le n\right\} \tag{5.13}$$

where $0 < t_1 < \cdots < t_n$ and $A_1, \ldots, A_n$ are Borel subsets of $\mathbb{R}$. We may define $W : \mathcal{F} \to [0,1]$ as a probability measure on the σ-algebra $\mathcal{F}$ on $C[0,\infty)$ that is generated by the cylinder sets.

It follows from (5.8) that the Wiener measure of the set (5.13) is given by

$$W\{\mathcal{C}\} = C_n \int_A \exp\left[-\frac{1}{2}\left\{\frac{(x_n - x_{n-1})^2}{(t_n - t_{n-1})} + \cdots + \frac{(x_1 - x_0)^2}{(t_1 - t_0)}\right\}\right] dx_1 dx_2 \ldots dx_n$$

where $A = A_1 \times A_2 \times \cdots \times A_n \subset \mathbb{R}^n$, $x_0 = 0$, $t_0 = 0$, and

$$C_n = \frac{1}{\sqrt{2\pi(t_n - t_{n-1})\ldots(t_1 - t_0))}}.$$

If we suppose, for simplicity, that $t_i - t_{i-1} = \Delta t$, then we may write this expression as

$$W\{\mathcal{C}\} = C_n \int_A \exp\left[-\frac{\Delta t}{2}\left\{\left(\frac{x_n - x_{n-1}}{\Delta t}\right)^2 + \cdots + \left(\frac{x_1 - x_0}{\Delta t}\right)^2\right\}\right] dx_1 dx_2 \ldots dx_n$$

Thus, formally taking the limit as $n \to \infty$, we get the expression given in (3.89)

$$dW = C \exp\left[-\frac{1}{2}\int_0^t \dot{x}^2(s)\, ds\right] Dx \tag{5.14}$$

for the density of Wiener measure with respect to the (unfortunately nonexistent) 'flat' measure Dx. Note that, since Wiener measure is supported on the set of continuous functions that are nowhere differentiable, the exponential factor in (5.14) makes no more sense than the 'flat' measure.

It is possible interpret (5.14) as defining a Gaussian measure in an infinite dimensional Hilbert space, but we will not consider that theory here. Instead, we will describe some properties of Wiener measure suggested by (5.14) that are, in fact, true despite the formal nature of the expression.

First, as we saw in Section 14.3, Kac's version of the Feynman-Kac formula is suggested by (5.14). Although it is difficult to make sense of Feynman's expression for solutions of the Schrödinger equation as an oscillatory path integral, Kac's formula for the heat equation with a potential makes perfect sense as an integral with respect to Wiener measure.

Second, (5.14) suggests the *Cameron-Martin theorem*, which states that the translation $x(t) \mapsto x(t) + h(t)$ maps Wiener measure W to a measure W_h that is absolutely continuous with respect to Wiener measure if and only if $h \in H^1(0,t)$ has a square integrable derivative. A formal calculation based on (5.14), and the idea that, like Lebesgue measure, Dx should be invariant under translations gives

$$\begin{aligned} dW_h &= C \exp\left[-\frac{1}{2}\int_0^t \left\{\dot{x}(s) - \dot{h}(s)\right\}^2 \, ds\right] \, Dx \\ &= C \exp\left[\int_0^t \dot{x}(s)\dot{h}(s)\, ds - \frac{1}{2}\int_0^t \dot{h}^2(s)\, ds\right] \exp\left[-\frac{1}{2}\int_0^t \dot{x}^2(s)\, ds\right] \, Dx \\ &= \exp\left[\int_0^t \dot{x}(s)\dot{h}(s)\, ds - \frac{1}{2}\int_0^t \dot{h}^2(s)\, ds\right] \, dW. \end{aligned}$$

The integral

$$\langle x, h\rangle = \int_0^t \dot{x}(s)\dot{h}(s)\, ds = \int_0^t \dot{h}(s)\, dx(s)$$

may be defined as a Payley-Wiener-Zygmund integral (5.47) for any $h \in H^1$. We then get the Cameron-Martin formula

$$dW_h = \exp\left[\langle x, h\rangle - \frac{1}{2}\int_0^t \dot{h}^2(s)\, ds\right] \, dW. \tag{5.15}$$

Despite the formal nature of the computation, the result is correct.

Thus, although Wiener measure is not translation invariant (which is impossible for probability measures on infinite-dimensional linear spaces) it is 'almost' translation invariant in the sense that translations in a dense set of directions $h \in H^1$ give measures that are mutually absolutely continuous. On the other hand, if one translates Wiener measure by a function $h \notin H^1$, one gets a measure that is singular with respect to the original Wiener measure, and which is supported on a set of paths with different continuity and variation properties.

These results reflect the fact that Gaussian measures on infinite-dimensional spaces are concentrated on a dense set of directions, unlike the picture we have of a finite dimensional Gaussian measure with an invertible covariance matrice (whose density is spread out over an ellipsoid in all direction).

4. Brownian motion with drift

Brownian motion is a basic building block for the construction of a large class of Markov processes with continuous sample paths, called diffusion processes.

In this section, we discuss diffusion processes that have the same 'noise' as standard Brownian motion, but differ from it by a mean 'drift.' These process are defined by a stochastic ordinary differential equation (SDE) of the form

$$\dot{X} = b(X) + \xi(t), \tag{5.16}$$

where $b : \mathbb{R} \to \mathbb{R}$ is a given smooth function and $\xi(t) = \dot{B}(t)$ is, formally, the time derivative of Brownian motion, or 'white noise.' Equation (5.16) may be thought of as describing either a Brownian motion $\dot{X} = \xi$ perturbed by a drift term $b(X)$, or a deterministic ODE $\dot{X} = b(X)$ perturbed by an additive noise.

We begin with a heuristic discussion of white noise, and then explain more precisely what meaning we give to (5.16).

4.1. White noise

Although Brownian paths are not differentiable pointwise, we may interpret their time derivative in a distributional sense to get a generalized stochastic process called white noise. We denote it by

$$\xi(t, \omega) = \dot{B}(t, \omega).$$

We also use the notation $\xi dt = dB$. The term 'white noise' arises from the spectral theory of stationary random processes, according to which white noise has a 'flat' power spectrum that is uniformly distributed over all frequencies (like white light). This can be observed from the Fourier representation of Brownian motion in (5.7), where a formal term-by-term differentiation yields a Fourier series all of whose coefficients are Gaussian random variables with same variance.

Since Brownian motion has Gaussian independent increments with mean zero, its time derivative is a Gaussian stochastic process with mean zero whose values at different times are independent. (See Figure 2.) As a result, we expect the SDE (5.16) to define a Markov process X. This process is not Gaussian unless $b(X)$ is linear, since nonlinear functions of Gaussian variables are not Gaussian.

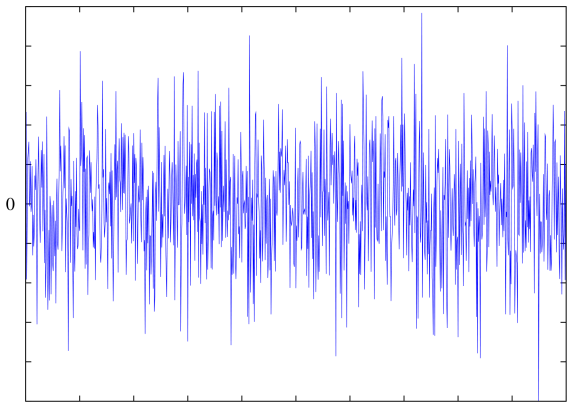

FIGURE 2. A numerical realization of an approximation to white noise.

To make this discussion more explicit, consider a finite difference approximation of ξ using a time interval of width Δt,

$$\xi_{\Delta t}(t) = \frac{B(t+\Delta t) - B(t)}{\Delta t}.$$

Then $\xi_{\Delta t}$ is a Gaussian stochastic process with mean zero and variance $1/\Delta t$. Using (5.11), we compute that its covariance is given by

$$\mathbf{E}\left[\xi_{\Delta t}(t)\xi_{\Delta t}(s)\right] = \delta_{\Delta t}(t-s)$$

where $\delta_{\Delta t}(t)$ is an approximation of the δ-function given by

$$\delta_{\Delta t}(t) = \frac{1}{\Delta t}\left(1 - \frac{|t|}{\Delta t}\right) \quad \text{if } |t| \leq \Delta t, \qquad \delta_{\Delta t}(t) = 0 \quad \text{otherwise.}$$

Thus, $\xi_{\Delta t}$ has a small but nonzero correlation time. Its power spectrum, which is the Fourier transform of its covariance, is therefore not flat, but decays at sufficiently high frequencies. We therefore sometimes refer to $\xi_{\Delta t}$ as 'colored noise.'

We may think of white noise ξ as the limit of this colored noise $\xi_{\Delta t}$ as $\Delta t \to 0$, namely as a δ-correlated stationary, Gaussian process with mean zero and covariance

$$\mathbf{E}\left[\xi(t)\xi(s)\right] = \delta(t-s). \tag{5.17}$$

In applications, the assumption of white noise is useful for modeling phenomena in which the correlation time of the noise is much shorter than any other time-scales of interest. For example, in the case of Brownian motion, the correlation time of the noise due to the impact of molecules on the Brownian particle is of the order of the collision time of the fluid molecules with each other. This is very small in comparison with the time-scales over which we use the SDE to model the motion of the particle.

4.2. Stochastic integral equations

While it is possible to define white noise as a distribution-valued stochastic process, we will not do so here. Instead, we will interpret white noise as a process whose time-integral is Brownian motion. Any differential equation that depends on white noise will be rewritten as an integral equation that depends on Brownian motion.

Thus, we rewrite (5.16) as the integral equation

$$X(t) = X(0) + \int_0^t b\left(X(s)\right)\, ds + B(t). \tag{5.18}$$

We use the differential notation

$$dX = b(X)dt + dB$$

as short-hand for the integral equation (5.18); it has no further meaning.

The standard Picard iteration from the theory of ODEs,

$$X_{n+1}(t) = X(0) + \int_0^t b\left(X_n(s)\right)\, ds + B(t),$$

implies that (5.18) has a unique continuous solution $X(t)$ for every continuous function $B(t)$, assuming that $b(x)$ is a Lipschitz-continuous function of x. Thus, if B is Brownian motion, the mapping $B(t) \mapsto X(t)$ obtained by solving (5.18) 'path by path' defines a stochastic process X with continuous sample paths. We call X a Brownian motion with drift.

Remark 5.16. According to Girsanov's theorem [**46**], the probability measure induced by X on $C[0,\infty)$ is absolutely continuous with respect to the Wiener measure induced by B, with density

$$\exp\left[\int_0^t b\left(X(s)\right)\, dX(s) - \frac{1}{2}\int_0^t b^2\left(X(s)\right)\, ds\right].$$

This is a result of the fact that the processes have the same 'noise,' so they are supported on the same paths; the drift changes only the probability density on those paths *c.f.* the Cameron-Martin formula (5.15).

4.3. The Fokker-Planck equation

We observed above that the transition density $p(x,t \mid y)$ of Brownian motion satisfies the diffusion equation (5.9). We will give a direct derivation of a generalization of this result for Brownian motion with drift.

We fix $y \in \mathbb{R}$ and write the conditional expectation given that $X(0) = y$ as

$$\mathbf{E}_y\left[\,\cdot\,\right] = \mathbf{E}\left[\,\cdot\,|X(0) = y\right].$$

Equation (5.18) defines a Markov process $X(t) = X_t$ with continuous paths. Moreover, as $\Delta t \to 0^+$, the increments of X satisfy

$$\mathbf{E}_y\left[X_{t+\Delta t} - X_t \mid X_t\right] = b\left(X_t\right)\Delta t + o\left(\Delta t\right), \tag{5.19}$$

$$\mathbf{E}_y\left[\left(X_{t+\Delta t} - X_t\right)^2 \mid X_t\right] = \Delta t + o\left(\Delta t\right), \tag{5.20}$$

$$\mathbf{E}_y\left[\left|X_{t+\Delta t} - X_t\right|^3 \mid X_t\right] = o(\Delta t), \tag{5.21}$$

where $o(\Delta t)$ denotes a term which approaches zero faster than Δt, meaning that

$$\lim_{\Delta t \to 0^+} \frac{o(\Delta t)}{\Delta t} = 0.$$

For example, to derive (5.19) we subtract (5.18) evaluated at $t + \Delta t$ from (5.18) evaluated at t to get

$$\Delta X = \int_t^{t+\Delta t} b\left(X_s\right)\, ds + \Delta B$$

where

$$\Delta X = X_{t+\Delta t} - X_t, \qquad \Delta B = B_{t+\Delta t} - B_t.$$

Using the smoothness of b and the continuity of X_t, we get

$$\begin{aligned}\Delta X &= \int_t^{t+\Delta t} \left[b\left(X_t\right) + o(1)\right]\, ds + \Delta B \\ &= b\left(X_t\right)\Delta t + \Delta B + o(\Delta t).\end{aligned}$$

Taking the expected value of this equation conditioned on X_t, using the fact that $\mathbf{E}[\Delta B] = 0$, and assuming we can exchange expectations with limits as $\Delta t \to 0^+$, we get (5.19). Similarly, Taylor expanding to second order, we find that the dominant term in $\mathbf{E}[(\Delta X)^2]$ is $\mathbf{E}[(\Delta B)^2] = \Delta t$, which gives (5.20). Equation (5.21) follows from the corresponding property of Brownian motion.

Now suppose that $\varphi : \mathbb{R} \to \mathbb{R}$ is any smooth test function with uniformly bounded derivatives, and let

$$e(t) = \frac{d}{dt}\mathbf{E}_y\left[\varphi\left(X_t\right)\right].$$

Expressing the expectation in terms of the transition density $p(x,t \mid y)$ of X_t, assuming that the time-derivative exists and that we may exchange the order of differentiation and expectation, we get

$$e(t) = \frac{d}{dt}\int \varphi(x)p(x,t \mid y)\,dx = \int \varphi(x)\frac{\partial p}{\partial t}(x,t \mid y)\,dx.$$

Alternatively, writing the time derivative as a limit of difference quotients, and Taylor expanding $\varphi(x)$ about $x = X_t$, we get

$$\begin{aligned} e(t) &= \lim_{\Delta t\to 0^+}\frac{1}{\Delta t}\mathbf{E}_y\left[\varphi\left(X_{t+\Delta t}\right)-\varphi\left(X_t\right)\right] \\ &= \lim_{\Delta t\to 0^+}\frac{1}{\Delta t}\mathbf{E}_y\left[\varphi'\left(X_t\right)\left(X_{t+\Delta t}-X_t\right)+\frac{1}{2}\varphi''\left(X_t\right)\left(X_{t+\Delta t}-X_t\right)^2+r_t(\Delta t)\right] \end{aligned}$$

where the remainder r_t satisfies

$$|r_t(\Delta t)| \le M\left|X_{t+\Delta t}-X_t\right|^3$$

for some constant M. Using the 'tower' property of conditional expectation (5.2) and (5.19), we have

$$\begin{aligned} \mathbf{E}_y\left[\varphi'\left(X_t\right)\left(X_{t+\Delta t}-X_t\right)\right] &= \mathbf{E}_y\left[\,\mathbf{E}_y\left[\varphi'\left(X_t\right)\left(X_{t+\Delta t}-X_t\right)\mid X_t\right]\right] \\ &= \mathbf{E}_y\left[\,\varphi'\left(X_t\right)\mathbf{E}_y\left[X_{t+\Delta t}-X_t\mid X_t\right]\right] \\ &= \mathbf{E}_y\left[\varphi'\left(X_t\right)b\left(X_t\right)\right]\Delta t. \end{aligned}$$

Similarly

$$\mathbf{E}_y\left[\,\varphi''\left(X_t\right)\left(X_{t+\Delta t}-X_t\right)^2\right] = \mathbf{E}_y\left[\,\varphi''\left(X_t\right)\right]\Delta t.$$

Hence,

$$e(t) = \mathbf{E}_y\left[\varphi'\left(X_t\right)b\left(X_t\right)+\frac{1}{2}\varphi''\left(X_t\right)\right].$$

Rewriting this expression in terms of the transition density, we get

$$e(t) = \int_{\mathbb{R}}\left[\varphi'(x)b(x)+\frac{1}{2}\varphi''(x)\right]p\left(x,t \mid y\right)\,dx.$$

Equating the two different expressions for $e(t)$ we find that,

$$\int_{\mathbb{R}}\varphi(x)\frac{\partial p}{\partial t}(x,t \mid y)\,dx = \int_{\mathbb{R}}\left[\varphi'(x)b(x)+\frac{1}{2}\varphi''(x)\right]p\left(x,t \mid y\right)\,dx.$$

This is the weak form of an advection-diffusion equation for the transition density $p(x,t \mid y)$ as a function of (x,t). After integrating by parts with respect to x, we find that, since φ is an arbitrary test function, smooth solutions p satisfy

$$\frac{\partial p}{\partial t} = -\frac{\partial}{\partial x}\left(bp\right)+\frac{1}{2}\frac{\partial^2 p}{\partial x^2}. \tag{5.22}$$

This PDE is called the Fokker-Planck, or forward Kolmogorov equation, for the diffusion process of Brownian motion with drift. When $b = 0$, we recover (5.9).

5. The Langevin equation

> A particle such as the one we are considering, large relative to the average distance between the molecules of the liquid and moving with respect to the latter at the speed ξ, experiences (according to Stokes' formula) a viscous resistance equal to $-6\pi\mu a\xi$. In actual fact, this value is only a mean, and by reason of the irregularity of the impacts of the surrounding molecules, the action of the fluid on the particle oscillates around the preceding value, to the effect that the equation of motion in the direction x is
>
> $$m\frac{d^2x}{dt^2} = -6\pi\mu a\frac{dx}{dt} + X.$$
>
> We know that the complementary force X is indifferently positive and negative and that its magnitude is such as to maintain the agitation of the particle, which, given the viscous resistance, would stop without it.[7]

In this section, we describe a one-dimensional model for the motion of a Brownian particle due to Langevin. A three-dimensional model may be obtained from the one-dimensional model by assuming that a spherical particle moves independently in each direction. For non-spherical particles, such as the pollen grains observed by Brown, rotational Brownian motion also occurs.

Suppose that a particle of mass m moves along a line, and is subject to two forces: (a) a frictional force that is proportional to its velocity; (b) a random white noise force. The first force models the average force exerted by a viscous fluid on a small particle moving though it; the second force models the fluctuations in the force about its mean value due to the impact of the fluid molecules.

This division could be questioned on the grounds that *all* of the forces on the particle, including the viscous force, ultimately arise from molecular impacts. One is then led to the question of how to derive a mesoscopic stochastic model from a more detailed kinetic model. Here, we will take the division of the force into a deterministic mean, given by macroscopic continuum laws, and a random fluctuating part as a basic hypothesis of the model. See Keizer [**30**] for further discussion of such questions.

We denote the velocity of the particle at time t by $V(t)$. Note that we consider the particle velocity here, not its position. We will consider the behavior of the position of the particle in Section 6. According to Newton's second law, the velocity satisfies the ODE

$$m\dot{V} = -\beta V + \gamma\xi(t), \tag{5.23}$$

where $\xi = \dot{B}$ is white noise, $\beta > 0$ is a damping constant, and γ is a constant that describes the strength of the noise. Dividing the equation by m, we get

$$\dot{V} = -bV + c\xi(t), \tag{5.24}$$

where $b = \beta/m > 0$ and $c = \gamma/m$ are constants. The parameter b is an inverse-time, so $[b] = T^{-1}$. Standard Brownian motion has dimension $T^{1/2}$ since $\mathbf{E}\left[B^2(t)\right] = t$, so white noise ξ has dimension $T^{-1/2}$, and therefore $[c] = LT^{-3/2}$.

---

[7] P. Langevin, *Comptes rendus Acad. Sci.* **146** (1908).

We suppose that the initial velocity of the particle is given by

$$V(0) = v_0, \tag{5.25}$$

where v_0 is a fixed deterministic quantity. We can obtain the solution for random initial data that is independent of the future evolution of the process by conditioning with respect to the initial value.

Equation (5.24) is called the *Langevin equation*. It describes the effect of noise on a scalar linear ODE whose solutions decay exponentially to the globally asymptotically stable equilibrium $V = 0$ in the absence of noise. Thus, it provides a basic model for the effect of noise on any system with an asymptotically stable equilibrium.

As explained in Section 4.2, we interpret (5.24)–(5.25) as an integral equation

$$V(t) = v_0 - b \int_0^t V(s)\, ds + cB(t), \tag{5.26}$$

which we write in differential notation as

$$dV = -bV dt + c dB.$$

The process $V(t)$ defined by (5.26) is called the *Ornstein-Uhlenbeck process*, or the OU process, for short.

We will solve this problem in a number of different ways, which illustrate different methods. In doing so, it is often convenient to use the formal properties of white noise; the correctness of any results we derive in this way can be verified directly.

One of the most important features of the solution is that, as $t \to \infty$, the process approaches a stationary process, called the stationary Ornstein-Uhlenbeck process. This corresponds physically to the approach of the Brownian particle to thermodynamic equilibrium in which the fluctuations caused by the noise balance the dissipation due to the damping terms. We will discuss the stationary OU process further in Section 6.

5.1. Averaging the equation

Since (5.24) is a linear equation for $V(t)$ with deterministic coefficients and an additive Gaussian forcing, the solution is also Gaussian. It is therefore determined by its mean and covariance. In this section, we compute these quantities by averaging the equation.

Let

$$\mu(t) = \mathbf{E}\left[V(t)\right].$$

Then, taking the expected value of (5.24), and using the fact that $\xi(t)$ has zero mean, we get

$$\dot{\mu} = -b\mu. \tag{5.27}$$

From (5.25), we have $\mu(0) = v_0$, so

$$\mu(t) = v_0 e^{-bt}. \tag{5.28}$$

Thus, the mean value of the process decays to zero in exactly the same way as the solution of the deterministic, damped ODE, $\dot{V} = -bV$.

Next, let

$$R\left(t, s\right) = \mathbf{E}\left[\left\{V(t) - \mu(t)\right\}\left\{V(s) - \mu(s)\right\}\right] \tag{5.29}$$

denote the covariance of the OU process. Then, assuming we may exchange the order of time-derivatives and expectations, and using (5.24) and (5.27), we compute that

$$\begin{aligned}\frac{\partial^2 R}{\partial t \partial s}(t,s) &= \mathbf{E}\left[\left\{\dot{V}(t) - \dot{\mu}(t)\right\}\left\{\dot{V}(s) - \dot{\mu}(s)\right\}\right] \\ &= \mathbf{E}\left[\{-b\left[V(t) - \mu(t)\right] + c\xi(t)\}\{-b\left[V(s) - \mu(s)\right] + c\xi(s)\}\right].\end{aligned}$$

Expanding the expectation in this equation and using (5.17), (5.29), we get

$$\frac{\partial^2 R}{\partial t \partial s} = b^2 R - bc\left\{L(t,s) + L(s,t)\right\} + c^2\delta(t-s) \tag{5.30}$$

where

$$L(t,s) = \mathbf{E}\left[\{V(t) - \mu(t)\}\xi(s)\right].$$

Thus, we also need to derive an equation for L. Note that $L(t,s)$ need not vanish when $t > s$ since then $V(t)$ depends on $\xi(s)$.

Using (5.24), (5.27), and (5.17), we find that

$$\begin{aligned}\frac{\partial L}{\partial t}(t,s) &= \mathbf{E}\left[\left\{\dot{V(t)} - \dot{\mu}(t)\right\}\xi(s)\right] \\ &= -b\mathbf{E}\left[\{V(t) - \mu(t)\}\xi(s)\right] + c\mathbf{E}\left[\xi(t)\xi(s)\right] \\ &= -bL(t,s) + c\delta(t-s).\end{aligned}$$

From the initial condition (5.25), we have

$$L(0,s) = 0 \qquad \text{for } s > 0.$$

The solution of this equation is

$$L(t,s) = \begin{cases} ce^{-b(t-s)} & \text{for } t > s, \\ 0 & \text{for } t < s. \end{cases} \tag{5.31}$$

This function solves the homogeneous equation for $t \neq s$, and jumps by c as t increases across $t = s$.

Using (5.31) in (5.30), we find that $R(t,s)$ satisfies the PDE

$$\frac{\partial^2 R}{\partial t \partial s} = b^2 R - bc^2 e^{-b|t-s|} + c^2\delta(t-s). \tag{5.32}$$

From the initial condition (5.25), we have

$$R(t,0) = 0, \quad R(0,s) = 0 \qquad \text{for } t, s > 0. \tag{5.33}$$

The second-order derivatives in (5.32) are the one-dimensional wave operator written in characteristic coordinates (t,s). Thus, (5.32)–(5.33) is a characteristic initial value problem for $R(t,s)$.

This problem has a simple explicit solution. To find it, we first look for a particular solution of the nonhomogeneous PDE (5.32). We observe that, since

$$\frac{\partial^2}{\partial t \partial s}\left(e^{-b|t-s|}\right) = -b^2 e^{-b|t-s|} + 2b\delta(t-s),$$

a solution is given by

$$R_p(t,s) = \frac{c^2}{2b}e^{-b|t-s|}.$$

Then, writing $R = R_p + \tilde{R}$, we find that $\tilde{R}(t,s)$ satisfies

$$\frac{\partial^2 \tilde{R}}{\partial t \partial s} = b^2\tilde{R}, \qquad \tilde{R}(t,0) = -\frac{c^2}{2b}e^{-bt}, \quad \tilde{R}(0,s) = -\frac{c^2}{2b}e^{-bs}.$$

This equation has the solution

$$\tilde{R}(t,s) = -\frac{c^2}{2b}\, e^{-b(t+s)}.$$

Thus, the covariance function (5.29) of the OU process defined by (5.24)–(5.25) is given by

$$R(t,s) = \frac{c^2}{2b}\left(e^{-b|t-s|} - e^{-b(t+s)}\right). \tag{5.34}$$

In particular, the variance of the process,

$$\sigma^2(t) = \mathbf{E}\left[\{V(t) - \mu(t)\}^2\right],$$

or $\sigma^2(t) = R(t,t)$, is given by

$$\sigma^2(t) = \frac{c^2}{2b}\left(1 - e^{-2bt}\right), \tag{5.35}$$

and the one-point probability density of the OU process is given by the Gaussian density

$$p(v,t) = \frac{1}{\sqrt{2\pi\sigma^2(t)}} \exp\left\{-\frac{[v-\mu(t)]^2}{2\sigma^2(t)}\right\}. \tag{5.36}$$

The success of the method used in this section depends on the fact that the Langevin equation is linear with additive noise. For nonlinear equations, or equations with multiplicative noise, one typically encounters the 'closure' problem, in which higher order moments appear in equations for lower order moments, leading to an infinite system of coupled equations for averaged quantities. In some problems, it may be possible to use a (more or less well-founded) approximation to truncate this infinite system to a finite system.

5.2. Exact solution

The SDE (5.24) is sufficiently simple that we can solve it exactly. A formal solution of (5.24) is

$$V(t) = v_0 e^{-bt} + c\int_0^t e^{-b(t-s)}\xi(s)\, ds. \tag{5.37}$$

Setting $\xi = \dot{B}$, and using a formal integration by parts, we may rewrite (5.37) as

$$V(t) = v_0 e^{-bt} + bc\left(B(t) - \int_0^t e^{-b(t-s)} B(s)\, ds\right). \tag{5.38}$$

This last expression does not involve any derivatives of $B(t)$, so it defines a continuous function $V(t)$ for any continuous Brownian sample function $B(t)$. One can verify by direct calculation that (5.38) is the solution of (5.26).

The random variable $V(t)$ defined by (5.38) is Gaussian. Its mean and covariance may be computed most easily from the formal expression (5.37), and they agree with the results of the previous section.

For example, using (5.37) in (5.29) and simplifying the result by the use of (5.17), we find that the covariance function is

$$\begin{aligned} R(t,s) &= c^2 \mathbf{E}\left[\left\{\int_0^t e^{-b\left(t-t'\right)}\xi\left(t'\right)\,dt'\right\}\left\{\int_0^s e^{-b\left(s-s'\right)}\xi\left(s'\right)\,ds'\right\}\right] \\ &= c^2\int_0^t\int_0^s e^{-b\left(t+s-t'-s'\right)}\mathbf{E}\left[\xi\left(t'\right)\xi\left(s'\right)\right]\,ds'dt' \\ &= c^2\int_0^t\int_0^s e^{-b\left(t+s-t'-s'\right)}\delta\left(t'-s'\right)\,ds'dt' \\ &= \frac{c^2}{2b}\left\{e^{-b|t-s|}-e^{-b(t+s)}\right\}. \end{aligned}$$

In more complicated problems, it is typically not possible to solve a stochastic equation exactly for each realization of the random coefficients that appear in it, so we cannot compute the statistical properties of the solution by averaging the exact solution. We may, however, be able to use perturbation methods or numerical simulations to obtain approximate solutions whose averages can be computed.

5.3. The Fokker-Planck equation

The final method we use to solve the Langevin equation is based on the Fokker-Planck equation. This method depends on a powerful and general connection between diffusion processes and parabolic PDEs.

From (5.22), the transition density $p(v,t \mid w)$ of the Langevin equation (5.24) satisfies the diffusion equation

$$\frac{\partial p}{\partial t} = \frac{\partial}{\partial v}\left(bvp\right) + \frac{1}{2}c^2\frac{\partial^2 p}{\partial v^2}. \tag{5.39}$$

Note that the coefficient of the diffusion term is proportional to c^2 since the Brownian motion cB associated with the white noise $c\xi$ has quadratic variation $\mathbf{E}\left[(c\Delta B)^2\right] = c^2\Delta t$.

To solve (5.39), we write it in characteristic coordinates associated with the advection term. (An alternative method is to Fourier transform the equation with respect to v, which leads to a first-order PDE for the transform since the variable coefficient term involves only multiplication by v. This PDE can then be solved by the method of characteristics.)

The sub-characteristics of (5.39) are defined by

$$\frac{dv}{dt} = -bv,$$

whose solution is $v = \tilde{v}e^{-bt}$. Making the change of variables $v \mapsto \tilde{v}$ in (5.39), we get

$$\frac{\partial p}{\partial t} = bp + \frac{1}{2}c^2 e^{2bt}\frac{\partial^2 p}{\partial \tilde{v}^2},$$

which we may write as

$$\frac{\partial}{\partial t}\left(e^{-bt}p\right) = \frac{1}{2}c^2 e^{2bt}\frac{\partial^2}{\partial \tilde{v}^2}\left(e^{-bt}p\right).$$

To simplify this equation further, we define

$$\tilde{p} = e^{-bt}p, \qquad \tilde{t} = \frac{c^2}{2b}\left(e^{2bt}-1\right),$$

which gives the standard diffusion equation

$$\frac{\partial \tilde{p}}{\partial \tilde{t}} = \frac{1}{2}\frac{\partial^2 \tilde{p}}{\partial \tilde{v}^2}.$$

The solution with initial condition

$$\tilde{p}\left(\tilde{v}, 0\right) = \delta\left(\tilde{v} - v_0\right)$$

is given by

$$\tilde{p}\left(\tilde{v}, \tilde{t}\right) = \frac{1}{(2\pi\tilde{t})^{1/2}}\, e^{-(\tilde{v}-v_0)^2/(2\tilde{t})}.$$

Rewriting this expression in terms of the original variables, we get (5.36).

The corresponding expression for the transition density is

$$p\left(v, t \mid v_0\right) = \frac{1}{\sqrt{2\pi\sigma^2(t)}} \exp\left\{-\frac{\left[v - v_0 e^{-bt}\right]^2}{2\sigma^2(t)}\right\}$$

where σ is given in (5.35).

Remark 5.17. It is interesting to note that the Ornstein-Uhlenbeck process is closely related to the 'imaginary' time version of the quantum mechanical simple harmonic oscillator. The change of variable

$$p(v,t) = \exp\left(\frac{1}{2}bx^2 - bt\right)\psi(x,t) \qquad v = cx,$$

transforms (5.39) to the diffusion equation with a quadratic potential

$$\frac{\partial \psi}{\partial t} = \frac{1}{2}\frac{\partial^2 \psi}{\partial x^2} - \frac{1}{2}b^2x^2\psi.$$

6. The stationary Ornstein-Uhlenbeck process

As $t \to \infty$, the Ornstein-Uhlenbeck process approaches a stationary Gaussian process with zero mean, called the stationary Ornstein-Uhlenbeck process. This approach occurs on a time-scale of the order b^{-1}, which is the time-scale for solutions of the deterministic equation $\dot{V} = -bV$ to decay to zero.

From (5.36) and (5.35), the limiting probability density for v is a Maxwellian distribution,

$$p(v) = \frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-v^2/(2\sigma^2)} \tag{5.40}$$

with variance

$$\sigma^2 = \frac{c^2}{2b}. \tag{5.41}$$

We can also obtain (5.40) by solving the ODE for steady solutions of (5.39)

$$\frac{1}{2}c^2\frac{d^2p}{dv^2} + b\frac{d}{dv}\left(vp\right) = 0.$$

Unlike Brownian paths, whose fluctatations grow with time, the stationary OU paths consist of fluctuations that are typically of the order σ, although larger fluctuations occur over long enough times.

The stationary OU process is the exact solution of the SDE (5.24) if, instead of taking deterministic initial conditions, we suppose that $V(0)$ is a Gaussian random variable with the stationary distribution (5.40).

Taking the limit as $t \to \infty$ in (5.34), we find that the covariance function of the stationary OU process is

$$R(t-s) = \sigma^2 \, e^{-b|t-s|}. \tag{5.42}$$

The covariance function depends only on the time-difference since the process is stationary. Equation (5.42) shows that the values of the stationary OU process become uncorrelated on the damping time-scale b^{-1}.

6.1. Parameter values for Brownian motion

Before we use use the OU process to determine the spatial diffusion of a Brownian particle, we give some typical experimental parameters for Brownian motion [**38**] and discuss their implications.

A typical radius of a spherical Brownian particle in water (for example, a polystyrene microsphere) is $a = 10^{-6}\,\mathrm{m}$. Assuming that the density of the particle is close to the density of water, its mass is approximately $m = 4 \times 10^{-15}\,\mathrm{Kg}$. According to Stokes law (2.24), at low Reynolds numbers, the viscous drag on a sphere of radius a moving with velocity v through a fluid with viscosity μ is equal to $6\pi\mu a v$. Thus, in (5.23), we take

$$\beta = 6\pi\mu a.$$

The viscosity of water at standard conditions is approximately $\mu = 10^{-3}\,\mathrm{Kg\,m^{-1}s^{-1}}$, which gives $\beta = 2 \times 10^{-8}\,\mathrm{Kg\,s^{-1}}$.

The first conclusion from these figures is that the damping time,

$$\frac{1}{b} = \frac{m}{\beta} \approx 2 \times 10^{-7}\,\mathrm{s},$$

is very small compared with the observation times of Brownian motion, which are typically on the order of seconds. Thus, we can assume that the Brownian particle velocity is in thermodynamic equilibrium and is distributed according to the stationary OU distribution. It also follows that the stationary OU fluctuations are very fast compared with the time scales of observation.

Although b^{-1} is small compared with macroscopic time-scales, it is large compared with molecular time scales; the time for water molecules to collide with each other is of the order of $10^{-11}\,\mathrm{s}$ or less. Thus, it is appropriate to use white noise to model the effect of fluctuations in the molecular impacts.

We can determine the strength of the noise in (5.23) by an indirect argument. According to statistical mechanics, the equilibrium probability density of a Brownian particle is proportional to $\exp(-E/kT)$, where $E = \frac{1}{2}mv^2$ is the kinetic energy of the particle, k is Boltzmann's constant, and T is the absolute temperature. This agrees with (5.40) if

$$\sigma^2 = \frac{kT}{m}. \tag{5.43}$$

At standard conditions, we have $kT = 4 \times 10^{-21}\,\mathrm{J}$, which gives $\sigma = 10^{-3}\,\mathrm{ms^{-1}}$. This is the order of magnitude of the thermal velocity fluctuations of the particle. The corresponding Reynolds numbers $\mathrm{R} = Ua/\nu$ are of the order 10^{-3} which is consistent with the use of Stokes' law.

Remark 5.18. It follows from (5.41) and (5.43) that

$$\gamma^2 = 2kT\beta.$$

This equation is an example of a *fluctuation-dissipation theorem*. It relates the macroscopic damping coefficient β in (5.23) to the strength γ^2 of the fluctuations when the system in thermodynamic equilibrium at temperature T.

6.2. The spatial diffusion of Brownian particles

Let us apply these results to the spatial diffusion of Brownian particles. We assume that the particles are sufficiently dilute that we can neglect any interactions between them.

Let $X(t)$ be the position at time t of a particle in Brownian motion measured along some coordinate axis. We assume that its velocity $V(t)$ satisfies the Langevin equation (5.23). Having solved for $V(t)$, we can obtain X by an integration

$$X(t) = \int_0^t V(s)\, ds.$$

Since $X(t)$ is a linear function of the Gaussian process $V(t)$, it is also Gaussian. The stochastic properties of X may be determined exactly from those of V, for example by averaging this equation to find its mean and covariance. We can, however, simplify the calculation when the parameters have the order of magnitude of the experimental ones given above.

On the time-scales over which we want to observe $X(t)$, the velocity $V(t)$ is a rapidly fluctuating, stationary Gaussian process with zero mean and a very short correlation time b^{-1}. We may therefore approximate it by white noise. From (5.42), the covariance function $R(t-s) = \mathbf{E}\,[V(t)V(s)]$ of V is given by

$$\mathbf{E}\,[V(t)V(s)] = \frac{2\sigma^2}{b}\left(\frac{be^{-b|t-s|}}{2}\right)$$

As $b \to \infty$, we have $be^{-b|t|}/2 \rightharpoonup \delta(t)$. Thus, from (5.17), if $bt \gg 1$, we may make the approximation

$$V(t) = \sqrt{\frac{2\sigma^2}{b}}\,\xi(t)$$

where $\xi(t)$ is a standard white noise.

It then follows that the integral of $V(t)$ is given in terms of a standard Brownian motion $B(t)$ by

$$X(t) = \sqrt{\frac{2\sigma^2}{b}}\,B(t).$$

The probability distribution of $X(t)$, which we denote $p(x,t)$, therefore satisfies the diffusion equation

$$\frac{\partial p}{\partial t} = D\frac{\partial^2 p}{\partial^2 x}$$

where, $D = \sigma^2/b$, or by use of (5.43),

$$D = \frac{kT}{\beta} \tag{5.44}$$

This is the result derived by Einstein (1905).

As Einstein observed, one can use (5.44) to determine Avogadro's number N_A, the number of molecules in one mole of gas, by measuring the diffusivity of Brownian

particles. Boltzmann's constant k is related to the macroscopically measurable gas constant R by $R = kN_A$; at standard conditions, we have $RT \approx 2,400\,\mathrm{J}$. Thus,

$$N_A = \frac{RT}{\beta D}$$

For the experimental values given above, with $\beta = 2 \times 10^{-8}\,\mathrm{Kg\,s^{-1}}$, the diffusivity of Brownian particles is found to be approximately $2 \times 10^{-13}\,\mathrm{m^2\,s^{-1}}$, meaning that the particles diffuse a distance on the order of a micron over a few seconds [**38**]. This gives $N_A \approx 6 \times 10^{23}$, consistent with the accepted value of $N_A = 6.02214 \times 10^{23}$, measured more accurately by other methods.

7. Stochastic differential equations

In this section, we discuss SDEs that are driven by white noise whose strength depends on the solution. Our aim here is to introduce some of the main ideas, rather than give a full discussion, and we continue to consider scalar SDEs. The ideas generalize to systems of SDEs, as we briefly explain in Section 7.5. For a more detailed introduction to the theory of SDEs, see [**19**]. For the numerical solution of SDEs, see [**32**]

The SDE (5.16) considered in Section 4 contains white noise with a constant strength. If the strength of the white noise depends on the solution, we get an SDE of the form

$$\dot{X} = b(X,t) + \sigma(X,t)\xi(t), \tag{5.45}$$

where $b, \sigma : \mathbb{R} \times [0,\infty) \to \mathbb{R}$ are smooth coefficient functions, which describe the drift and diffusion, respectively. We allow the coefficients to depend explicitly on t.

As we will explain, there is a fundamental ambiguity in how to interpret an SDE such as (5.45) which does not arise when σ is constant.

First, we rewrite (5.45) as an integral equation for $X(t)$,

$$X(t) = X(0) + \int_0^t b\left(X(s), s\right)\, ds + \int_0^t \sigma\left(X(s), s\right) dB(s), \tag{5.46}$$

or, in differential notation, as

$$dX = b(X,t)\, dt + \sigma(X,t)\, dB.$$

We interpret (5.45) as the corresponding integral equation (5.46). In order to do so, we need to define the stochastic integral

$$\int_0^t \sigma\left(X(s), s\right) dB(s).$$

When $\sigma = 1$, we made the obvious definition that this integral is to equal $B(t)$. More generally, if $F(t)$ is a stochastic process with smooth sample paths, we can define the integral of F against dB by use of a formal integration by parts:

$$\int_0^t F(s) dB(s) = F(t)B(t) - \int_0^t \dot{F}(s)B(s)\, ds.$$

For deterministic integrands, we can relax the smoothness condition and define a stochastic integral for any $f \in L^2(0,t)$ such that

$$\int_0^t f^2(s)\, ds < \infty.$$

If $f(s)$ is smooth and $f(t) = 0$, then (by a formal white-noise computation, which is easy to verify [**19**])

$$\begin{aligned}\mathbf{E}\left[\left\{\int_0^t f(s)dB(s)\right\}^2\right] &= \int_0^t \int_0^t f(s)f(r)\mathbf{E}\left[\xi(s)\xi(r)\right]\, dsdr \\ &= \int_0^t \int_0^t f(s)f(r)\delta(s-r)\, dsdr \\ &= \int_0^t f^2(s)\, ds.\end{aligned}$$

If $f \in L^2(0,t)$, then we choose a sequence of smooth functions f_n such that $f_n \to f$ with respect to the L^2-norm. We then define

$$\int_0^t f(s)dB(s) = \lim_{n\to\infty} \int_0^t f_n(s)dB(s), \tag{5.47}$$

where, from the preceding estimate, the integrals converge in the sense of mean-square expectation,

$$\lim_{n\to\infty} \mathbf{E}\left[\left\{\int_0^t f_n(s)dB(s) - \int_0^t f(s)dB(s)\right\}^2\right] = 0.$$

This definition of a stochastic integral is due to Payley, Wiener, and Zygmund (1933).

None of these definitions work, however, if F is a stochastic process with continuous but non-differentiable paths, such as a function of B or X of (5.46), which is exactly the case we are interested in.

In the next section, we illustrate the difficulties that arise for such integrals. We will then indicate how to define the Itô integral, which includes the above definitions as special cases.

7.1. An illustrative stochastic integral

Let $B(t)$ be a standard Brownian motion starting at 0. Consider, as a specific example, the question of how to define the integral

$$J(t) = \int_0^t B(s)\, dB(s) \tag{5.48}$$

by the use of Riemann sums. We will give two different definitions, corresponding to the Strantonovich and Itô integral, respectively.

Let $0 = s_0 < s_1 < \cdots < s_n < s_{n+1} = t$ be a non-random partition of $[0,t]$. The Strantonovich definition of (5.48) corresponds to a limit of centered Riemann sums, such as

$$J_n^{(\mathrm{S})} = \sum_{i=0}^{n} \frac{1}{2}\left[B\left(s_{i+1}\right) + B\left(s_i\right)\right]\left[B\left(s_{i+1}\right) - B\left(s_i\right)\right].$$

This gives a telescoping series with the sum

$$J_n^{(\mathrm{S})} = \frac{1}{2}\sum_{i=0}^{n}\left[B^2\left(s_{i+1}\right) - B^2\left(s_i\right)\right] = \frac{1}{2}\left[B^2\left(s_{n+1}\right) - B^2\left(s_0\right)\right].$$

Thus, we get the Strantonovich integral

$$\int_0^{t\,(\mathrm{S})} B(s)\,dB(s) = \frac{1}{2}B^2(t), \tag{5.49}$$

as in the usual calculus. The Strantonovich definition of the integral is, however, not well-suited to the Markov and martingale properties of stochastic processes. For example, the expected value of the Strantonovich integral in (5.49) is nonzero and equal to $t/2$.

The Itô definition of (5.48) corresponds to a limit of forward-differenced Riemann sums, such as

$$J_n^{(\mathrm{I})} = \sum_{i=0}^{n} B\left(s_i\right)\left[B\left(s_{i+1}\right) - B\left(s_i\right)\right].$$

We can rewrite this equation as

$$\begin{aligned} J_n^{(\mathrm{I})} &= \frac{1}{2}\sum_{i=0}^{n}\left[\left\{B\left(s_{i+1}\right) + B\left(s_i\right)\right\} - \left\{B\left(s_{i+1}\right) - B_i\left(s_i\right)\right\}\right]\left[B\left(s_{i+1}\right) - B\left(s_i\right)\right] \\ &= \frac{1}{2}\sum_{i=0}^{n}\left[B^2\left(s_{i+1}\right) - B^2\left(s_i\right)\right] - \frac{1}{2}\sum_{i=0}^{n}\left[B\left(s_{i+1}\right) - B\left(s_i\right)\right]^2. \end{aligned}$$

The first sum gives $B^2(t)$, as for the Strantonovich integral, while the second sum converges almost surely to t as $n \to \infty$ by the quadratic-variation property of Brownian motion.

The Itô integral is therefore

$$\int_0^{t\,(\mathrm{I})} B(s)\,dB(s) = \frac{1}{2}\left[B^2(t) - t\right]. \tag{5.50}$$

This definition has powerful stochastic properties; for example, it defines a martingale, consistent with the fact that the expected value of the Itô integral in (5.50) is equal to zero.

If we use the Itô definition, however, the usual rules of calculus must be modified to include (5.12). For example, the differential form of (5.50) may be derived formally as follows:

$$d\left(\frac{1}{2}B^2\right) = \frac{1}{2}\left[(B + dB)^2 - B^2\right] = BdB + \frac{1}{2}(dB)^2 = BdB + \frac{1}{2}dt. \tag{5.51}$$

As this example illustrates, there is an inherent ambiguity in how one defines stochastic integrals such as (5.48). This ambiguity is caused by the sensitivity of the values of the Riemann sums to the location of the point where one evaluates the integrand, which is a result of the unbounded total variation of the Brownian sample paths.

We will use the Itô definition, but it should be emphasized that this choice is a matter of mathematical convenience. For instance, one can express the Itô and Strantonovich integrals in terms of each other.

7.2. The Itô integral

We will not define the Itô's integral in detail, but we will give a brief summary of some of the main points. Evans [**19**] or Varadhan [**46**] give proofs of most of the results stated here.

A stochastic process

$$F : [0, \infty) \times \Omega \to \mathbb{R}$$

is said to be *adapted* to a Brownian motion $B(t)$ if, for each $t \geq 0$, $F(t)$ is measurable with respect to the σ-algebra $\mathcal{F}_t$ generated by the random variables $\{B(s) : 0 \leq s \leq t\}$. Roughly speaking, this means that $F(t)$ is a function of $\{B(s) : 0 \leq s \leq t\}$.

If $F(t)$ is an adapted process with almost surely continuous sample paths and

$$\int_0^t \mathbf{E}\left[F^2(s)\right]\, ds < \infty,$$

then we can define the stochastic Itô integral of F with respect to B as a limit in mean-square expectation of forward-differenced Riemann sums

$$\int_0^t F(s)\, dB(s) = \lim_{n\to\infty} \sum_{i=0}^{n} F\left(s_i\right)\left[B\left(s_{i+1}\right) - B\left(s_i\right)\right],$$

or, in general, as a limit of integrals of adapted simple functions.

An important property of the Itô integral is that, as in (5.50),

$$\mathbf{E}\left[\int_0^t F(s)\, dB(s)\right] = 0. \tag{5.52}$$

This follows because $F(t)$ is independent of $B(t + \Delta t) - B(t)$ for $\Delta t > 0$, since F is adapted, so

$$\mathbf{E}\left[F\left(s_i\right)\left\{B\left(s_{i+1}\right) - B\left(s_i\right)\right\}\right] = \mathbf{E}\left[F\left(s_i\right)\right]\mathbf{E}\left[B\left(s_{i+1}\right) - B\left(s_i\right)\right] = 0.$$

Since Brownian motion has independent increments, one can see by a similar argument that the Itô integral

$$M(t) = M_0 + \int_0^t F(s)\, dB(s) \tag{5.53}$$

defines a martingale, meaning that $\mathbf{E}\left[M(t) \mid \mathcal{F}_s\right] = M(s)$ for $0 \leq s < t$.

We then define the Itô SDE

$$dX = b(X, t)\, dt + \sigma(X, t)\, dB \tag{5.54}$$

by (5.46), where the integral is understood to be an Itô integral. The initial data

$$X(0) = X_0 \tag{5.55}$$

is a given $\mathcal{F}_0$-measurable random variable. Here, we allow the initial value $B(0)$ of the Brownian motion to be a random variable, and $\mathcal{F}_0 = \sigma\left(B(0)\right)$.

For the SDE (5.18) with constant noise, we can define solutions 'path by path.' For (5.46), the definition of a solution depends on a probabilistic convergence of the integral. Thus, it is essentially stochastic in nature.

It can be shown that the SDE (5.54)–(5.55) has a unique adapted solution $X(t)$ with continuous paths defined for all $0 \leq t \leq T$ if, for example:

(1) the functions $b, \sigma : \mathbb{R} \times [0, T] \to \mathbb{R}$ are continuous, globally Lipschitz in x, and uniformly bounded in t, meaning that there exists a constant K such that for all $x, y \in \mathbb{R}$, $t \in [0, T]$

$$|b(x,t) - b(y,t)| \leq K|x - y|, \quad |\sigma(x,t) - \sigma(y,t)| \leq K|x - y|,$$
$$|b(x,t)| \leq K\left(1 + |x|\right), \quad |\sigma(x,t)| \leq K\left(1 + |x|\right);$$

(2) the initial data satisfies

$$\mathbf{E}\left[X_0^2\right] < \infty. \tag{5.56}$$

Such solutions are called strong solutions. It is also possible to define weak solutions of that satisfy the SDE in a distributional sense, and which exist even if the coefficient functions are not Lipschitz continuous, but we will not use weak solutions here.

7.3. Itô's formula

As we saw above, it is necessary to modify the usual rules of calculus if one uses Itô integrals. The key result is a version of the chain rule called Itô's formula.

Suppose that $X(t)$ is a solution of the Itô SDE (5.54), and $f(X,t)$ is a smooth function $f : \mathbb{R} \times [0,\infty) \to \mathbb{R}$. Here, we abuse notation slightly and use the same symbol for the argument of f and the process. Define

$$Y(t) = f\left(X(t), t\right).$$

Then Itô's formula states that Y satisfies the SDE

$$\begin{aligned} dY = &\left[\frac{\partial f}{\partial t}(X,t) + b(X,t)\frac{\partial f}{\partial X}(X,t) + \frac{1}{2}\sigma^2(X,t)\frac{\partial^2 f}{\partial X^2}(X,t)\right] dt \\ &+ \sigma(X,t)\frac{\partial f}{\partial X}(X,t)\, dB. \end{aligned} \tag{5.57}$$

This equation stands, of course, for the corresponding stochastic integral equation. Equation (5.57) is what one would obtain from the usual chain rule with an additional term in the drift proportional to the second x-derivative of f. In particular, if $X = B$, then $b = 0$, $\sigma = 1$, and Itô's formula becomes

$$df(B,t) = \left[\frac{\partial f}{\partial t}(B,t) + \frac{1}{2}\frac{\partial f^2}{\partial B^2}(B,t)\right] dt + \frac{\partial f}{\partial B}(B,t)\, dB. \tag{5.58}$$

For a proof of (5.57), see [**19**].

Itô's formula may be motivated by a formal computation using (5.54) and (5.12). For example, when $Y = f(X)$ we get, denoting X-derivatives by primes,

$$\begin{aligned} dY &= f'(X)dX + \frac{1}{2}f''(X)dX^2 \\ &= f'(X)\left[b(X,t)\, dt + \sigma(X,t)\, dB\right] + \frac{1}{2}f''(X)\sigma^2(X,t)\ (dB)^2 \\ &= \left[f'(X)b(X,t) + \frac{1}{2}f''(X)\sigma^2(X,t)\right] dt + f'(X)\sigma(X,t)\, dB. \end{aligned}$$

Example 5.19. Itô's formula (5.58) gives, as in (5.51),

$$d\left(\frac{1}{2}B^2\right) = \frac{1}{2}dt + BdB.$$

Example 5.20. If $f(B) = e^{\sigma B}$, where σ is a constant, then (5.58) implies that

$$de^{\sigma B} = \frac{1}{2}\sigma^2 e^{\sigma B}\, dt + \sigma e^{\sigma B}\, dB.$$

Taking expected values of this equation, and using the martingale property (5.52) of the Itô integral, we find that

$$d\mathbf{E}\left[e^{\sigma B}\right] = \frac{1}{2}\sigma^2 \mathbf{E}\left[e^{\sigma B}\right].$$

Solving this equation, and assuming that $B(t)$ starts at 0, we find that

$$\mathbf{E}\left[e^{\sigma B}\right] = e^{\sigma^2 t/2}. \tag{5.59}$$

7.4. The Fokker-Planck equation

Itô's formula provides a quick and efficient way to derive the Fokker-Planck equation. Suppose that $X(t)$ satisfies

$$dX = b(X,t)\,dt + \sigma(X,t)\,dB.$$

Taking the expectation of Itô's formula (5.57) and using the martingale property (5.52), we find that for any smooth function $f : \mathbb{R} \to \mathbb{R}$,

$$\mathbf{E}\left[f\left(X(t)\right)\right] = \int_0^t \mathbf{E}\left[f'\left(X(s)\right) b\left(X(s),s\right) + \frac{1}{2} f''\left(X(s)\right) \sigma^2\left(X(s),s\right)\right] ds.$$

Differentiating this equation with respect to t, we get

$$\frac{d}{dt}\mathbf{E}\left[f\left(X(t)\right)\right] = \mathbf{E}\left[f'\left(X(t)\right) b\left(X(t),t\right) + \frac{1}{2} f''\left(X(t)\right) \sigma^2\left(X(t),t\right)\right].$$

Writing this equation in terms of the probability density $p(x,t)$, or the transition density $p(x,t \mid y,s)$ if we condition on $X(s) = y$, we get

$$\frac{d}{dt}\int f(x)\,p(x,t)\,dx = \int \left[f'(x)\,b(x,t) + \frac{1}{2} f''(x)\,\sigma^2(x,t)\right] p(x,t)\,dx,$$

which is the weak form of the Fokker-Planck equation,

$$\frac{\partial p}{\partial t} = -\frac{\partial}{\partial x}\left(b(x,t)\,p\right) + \frac{1}{2}\frac{\partial^2}{\partial^2 x}\left(\sigma^2(x,t)\,p\right). \tag{5.60}$$

7.5. Systems of SDEs

A system of SDEs for a vector-valued stochastic process $\vec{X}(t) = (X_1(t), \ldots, X_n(t))$ may be written as

$$d\vec{X} = \vec{b}\left(\vec{X},t\right)\,dt + \sigma\left(\vec{X},t\right)\,d\vec{B}. \tag{5.61}$$

In (5.61), the vector $\vec{B}(t) = (B_1(t), \ldots, B_n(t))$ is an n-dimensional Brownian motion whose components $B_i(t)$ are independent one-dimensional Brownian motions such that

$$\mathbf{E}\left[B_i(t)B_j(s)\right] = \begin{cases} \min(t,s) & \text{if } i = j, \\ 0 & \text{if } i \neq j. \end{cases}$$

The coefficient functions in (5.61) are a drift vector $\vec{b} = (b_1, \ldots, b_n)$ and a diffusion matrix $\sigma = (\sigma_{ij})$

$$\vec{b} : \mathbb{R}^n \times [0,\infty) \to \mathbb{R}^n, \qquad \sigma : \mathbb{R}^n \times [0,\infty) \to \mathbb{R}^{n\times n},$$

which we assume satisfy appropriate smoothness conditions.

The differential form of the SDE (5.61) is short-hand for the integral equation

$$\vec{X}(t) = \vec{X}_0 + \int_0^t \vec{b}\left(\vec{X}(s),s\right)\,ds + \int_0^t \sigma\left(\vec{X}(s),s\right)\,d\vec{B}(s),$$

or, in component form,

$$X_i(t) = X_{i0} + \int_0^t b_i\left(X_1(s), \ldots, X_n(s), s\right)\, ds + \sum_{j=1}^n \int_0^t \sigma_{ij}\left(X_1(s), \ldots, X_n(s), s\right)\, dB_j(s) \qquad \text{for } 1 \leq i \leq n.$$

The integrals here are understood as Itô integrals.

If $f : \mathbb{R}^n \times [0, \infty) \to \mathbb{R}$ is a smooth function $f(X_1, \ldots, X_n, t)$, and

$$Y(t) = f\left(X_1(t), \ldots, X_n(t), t\right)$$

where $\vec{X}(t) = (X_1(t), \ldots, X_n(t))$ is a solution of (5.61), then Itô's formula is

$$dY = \left(\frac{\partial f}{\partial t} + \sum_{i=1}^n b_i \frac{\partial f}{\partial X_i} + \frac{1}{2} \sum_{i,j,k=1}^n \sigma_{ik}\sigma_{jk} \frac{\partial^2 f}{\partial X_i \partial X_j} \right) dt + \sum_{i,j=1}^n \sigma_{ij} \frac{\partial f}{\partial X_i}\, dB_j. \tag{5.62}$$

This result follows formally from the generalization of (5.12) to the 'rule'

$$dB_i\, dB_j = \begin{cases} dt & \text{if } i = j, \\ 0 & \text{if } i \neq j. \end{cases}$$

The coefficients of the resulting drift terms in (5.62) are

$$a_{ij} = \sum_{k=1}^n \sigma_{ik}\sigma_{jk}.$$

Thus, $A = (a_{ij})$ is given by $A = \sigma\sigma^\top$.

The Fokker-Planck equation for the transition density $p\left(\vec{x}, t \mid \vec{y}, s\right)$ may be derived in the same way as in the scalar case. The result is that

$$\frac{\partial p}{\partial t} = -\sum_{i=1}^n \frac{\partial}{\partial x_i}\left(b_i p\right) + \frac{1}{2} \sum_{i,j=1}^n \frac{\partial^2}{\partial x_i \partial x_j}\left(a_{ij} p\right),$$

with the initial condition $p\left(\vec{x}, s \mid \vec{y}, s\right) = \delta\left(\vec{x} - \vec{y}\right)$.

7.6. Strantonovich SDEs

Suppose that X satisfies the Strantonovich SDE

$$dX = b(X, t)\, dt + \sigma(X, t) \circ dB \tag{5.63}$$

where the notation $\circ dB$ indicates that the corresponding integral in (5.46) is to be interpreted as a Strantonovich integral. Then the normal rules of calculus apply, and $Y = f(X)$ satisfies

$$dY = f'(X) b(X, t)\, dt + f'(X)\sigma(X, t) \circ dB.$$

The derivation of the Fokker-Planck equation is not as simple as for the Itô SDE, since the expected value of a Strantonovich integral is, in general, nonzero, but one can show that the Fokker-Planck equation for (5.63) is

$$\begin{aligned} \frac{\partial p}{\partial t} &= -\frac{\partial}{\partial x}\left(b\left(x, t\right) p\right) + \frac{1}{2} \frac{\partial}{\partial x}\left[\sigma(x, t) \frac{\partial}{\partial x}\left(\sigma\left(x, t\right) p\right)\right] \\ &= -\frac{\partial}{\partial x}\left\{\left[b\left(x, t\right) + \frac{1}{2}\sigma\left(x, t\right) \frac{\partial \sigma}{\partial x}\left(x, t\right)\right] p\right\} + \frac{1}{2} \frac{\partial^2}{\partial^2 x}\left[\sigma^2(x, t) p\right]. \end{aligned}$$

If σ is not constant, this PDE has a different drift term than the one in (5.60) arising from the Itô SDE (5.54).

Equivalently, the solution $X(t)$ of the Strantonovich SDE (5.63) is the same as the solution of the Itô SDE

$$dX = \left[b(X,t) + \frac{1}{2}\sigma(X,t)\frac{\partial \sigma}{\partial X}(X,t) \right] dt + \sigma(X)\, dB$$

with a corrected drift. Thus, the difference in drifts is simply a consequence of the difference in the definitions of the Itô and Strantonovich integrals, and it has no other significance. Of course, in using an SDE to model a system, one must choose an appropriate drift and noise, The drift will therefore depend on what definition of the stochastic integral one uses (see Remark 5.21).

8. Financial models

In this section we describe a basic SDE models of a financial market and derive the Black-Scholes formula for options pricing.

8.1. Stock prices

A simple model for the dynamics of the price $S(t) > 0$ of a stock at time t, introduced by Samuelson (1965), is provided by the Itô SDE

$$dS = \mu S\, dt + \sigma S\, dB \tag{5.64}$$

where μ and σ are constant parameters.

The drift-constant μ in (5.64) is the expected rate of return of the stock; in the absence of noise, $S(t) = S_0 e^{\mu t}$. The noise term describes random fluctuations in the stock price due to the actions of many individual investors. The strength of the noise is σS since we expect that the fluctuations in the price of a stock should be proportional to its price. The diffusion-constant σ is called the volatility of the stock; it is larger for more speculative stocks. Typical values for σ are in the range 0.2–0.4 in units of $(\text{years})^{1/2}$, corresponding to a standard deviation in the relative stock price of 20–40 percent per annum.

The dependence of the noise in (5.64) on the solution S differs from the constant noise in the Ornstein-Uhlenbeck SDE, which describes physical systems in thermodynamic equilibrium where the noise is fixed by the temperature.

As can be verified by the use of Itô's formula, the exact solution of (5.64) is

$$S(t) = S_0 \exp\left[\left(\mu - \frac{1}{2}\sigma^2 \right) t + \sigma B(t) \right] \tag{5.65}$$

where S_0 is the initial value of S. The process (5.65) is called *geometric Brownian motion*. The logarithm of $S(t)$ is Gaussian, meaning that $S(t)$ is lognormal. From (5.65) and (5.59), the expected value of $S(t)$ is

$$\mathbf{E}\left[S(t)\right] = \mathbf{E}\left[S_0\right] e^{\mu t},$$

consistent with what one obtains by averaging (5.64) directly.

Remark 5.21. We could equally well model the stock price by use of a Strantonovich SDE with a corrected value for the drift

$$dS = \left(\mu - \frac{1}{2}\sigma^2 \right) S\, dt + \sigma S\, \partial B. \tag{5.66}$$

The growth rate of the drift term in this equation is lower than the growth rate of the drift term in the corresponding Itô equation. This is because the Strantonovich noise contributes to the mean growth rate. Favorable fluctuations in the stock price increase the growth rate due to noise, and this outweighs the effect of unfavorable fluctuations that decrease the growth rate. The noise term in the Itô equation is defined so that its mean effect is zero. The solution of (5.66), which is found by the usual rules of calculus, is the same as the solution (5.65) of the corresponding Itô equation.

8.2. An ideal market

Consider, as an idealized model, a financial market that consists of single stock whose price $S(t)$ satisfies (5.64), and a risk-free security, such as a bond, whose price $R(t)$ satisfies the deterministic equation

$$dR = rR\,dt. \tag{5.67}$$

Thus, the value of the risk-free security is unaffected by random fluctuations in the stock market, and is assumed to have a fixed constant rate of return r.

We will refer to any item that is traded on the market, such as the stock, the bond, or a derivative, as a security. The prices, or values, of securities and the amounts owned by different traders are stochastic processes that are adapted to the filtration $\{\mathcal{F}_t : t \geq 0\}$ generated by the Brownian motion $B(t)$ in (5.64). This means that we cannot look into the future.

We assume that all processes have continuous sample paths. We further assume, for simplicity, that we can trade continuously without cost or restriction, that stocks and bonds are infinitely divisible, and that we can neglect any other complicating factors, such as dividends.

A portfolio is a collection of investments. If a portfolio consists of $a_i(t)$ units of securities with values $V_i(t)$, where $1 \leq i \leq n$, the value $\Pi(t)$ of the portfolio is

$$\Pi = \sum_{i=1}^{n} a_i V_i. \tag{5.68}$$

The value of the portfolio satisfies an SDE of the form

$$d\Pi = b\,dt + c\,dB.$$

We say that the portfolio is *risk-free* if $c = 0$, meaning that its value is not directly affected by random fluctuations in the market. Without further assumptions, however, the growth rate b could depend on B.

We say that the portfolio is *self-financing* if

$$d\Pi = \sum_{i=1}^{n} a_i\,dV_i. \tag{5.69}$$

As usual, this equation stands for the corresponding Itô integral equation. The condition (5.69) means that the change in the value of the portfolio is entirely due to the change in value of the securities it contains. Therefore, after the initial investment, no money flows in or out of the portfolio.

We will take as a basic assumption that the market allows no arbitrage opportunities in which traders can make a guaranteed profit through multiple transactions. Specifically, we assume that the value $\Pi(t)$ of any self-financing, risk-free security

must satisfy the ODE

$$d\Pi = r\Pi\, dt \tag{5.70}$$

where r is the risk-free rate of return in (5.67).

If there were a self-financing, risk-free portfolio whose instantaneous rate of return was higher (or lower) than the prevailing rate r, then traders could make a guaranteed profit by continuously buying (or selling) the securities in the portfolio. This would rapidly drive the rate of return of the portfolio and the prevailing rate r to the same value, which is the theoretical justification of the no-arbitrage assumption.

8.3. Derivatives

> It is a recipe for disaster to give one or two people complete authority to trade derivatives without a close monitoring of the risks being taken.[8]

Next, let us use this model to study the pricing of derivatives such as stock options. A derivative is a financial instrument that derives its value from some underlying asset. The asset could be almost anything, from pork bellies to next season's snowfall at a ski resort. Here, we consider derivatives that are contingent on the price of a stock.

We assume that the value $V(t)$ of the derivative is a deterministic function of the stock price $S(t)$ and the time t,

$$V(t) = f\left(S(t), t\right), \qquad f : (0,\infty) \times [0,\infty) \to \mathbb{R}.$$

Our aim is to determine what functions $f(S,t)$ provide values for a derivative that are consistent with the no-arbitrage assumption. The idea, following Black-Scholes (1973) and Merton (1973), is to construct a risk-free portfolio whose value replicates the value of the derivative.

Suppose that we sell, or write, one derivative, and form a portfolio that consists of:

(1) the derivative (whose value is a liability to us);
(2) a quantity $a(t)$ of the risk-free security with price $R(t)$;
(3) a quantity $b(t)$ of stock with price $S(t)$.

The value $\Pi(t)$ of the portfolio is given by

$$\Pi = aR + bS - V, \tag{5.71}$$

where R satisfies (5.67) and S satisfies (5.64).

We will choose $a(t)$, $b(t)$ so that the portfolio is self-financing and risk-free. In that case, its value must grow at the risk-free rate of return, and this will tell us how the value of the derivative $V(t)$ changes in terms of the price $S(t)$ of the stock.

The role of $R(t)$ is simply to provide a source of funds within the portfolio which allows us to adjust the stock holding as the value of the derivative fluctuates in order to maintain a risk-free position (this is called 'hedging'); R does not appear in the final result. If we did not include R in the portfolio, we would need to move funds in and out of the portfolio to make it risk-free.

From (5.69) and (5.71), the portfolio is self-financing if

$$d\Pi = a\, dR + b\, dS - dV. \tag{5.72}$$

[8] J. C. Hull, *Options Futures and Other Derivatives*, 4th ed., 2000.

Writing $V(t) = f\,(S(t), t)$, then using Itô's formula and (5.64), we get

$$dV = \left(\frac{\partial f}{\partial t} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 f}{\partial S^2}\right) dt + \frac{\partial f}{\partial S}\, dS.$$

Using this result and (5.67) in (5.72), we find that

$$d\Pi = \left(raR - \frac{\partial f}{\partial t} - \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 f}{\partial S^2}\right) dt + \left(b - \frac{\partial f}{\partial S}\right) dS.$$

Hence, the portfolio is risk-free if

$$b = \frac{\partial f}{\partial S}, \tag{5.73}$$

in which case

$$d\Pi = \left(raR - \frac{\partial f}{\partial t} - \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 f}{\partial S^2}\right) dt.$$

The no-arbitrage assumption (5.70) then implies that

$$d\Pi = r\Pi\, dt.$$

Equating these expressions for $d\Pi$, using (5.71) and (5.73), and simplifying the result, we find that

$$\frac{\partial f}{\partial t} + rS\frac{\partial f}{\partial S} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 f}{\partial S^2} = rf. \tag{5.74}$$

This is a PDE for $f(S, t)$, called the *Black-Scholes PDE*.

It is interesting to note that the rate of return μ of the stock in (5.64) does not appear in (5.74). The equation involves only the volatility σ of the stock and the risk-free rate of return r.

Equation (5.74) is a backward diffusion equation. In principle, any solution provides a feasible value-function for a derivative that is contingent on the stock. In the next section, we use (5.74) to determine the pricing of an option. The value of an option is known at the time when it comes to maturity, and this provides a final condition for (5.74). Solving the PDE backward in time then determines the initial price of the option. As we will see, although (5.74) has variable coefficients, we can obtain explicit solutions by transforming it to a constant coefficient heat equation.

8.4. Options

An option gives the holder the right, but not the obligation, to buy or sell an underlying asset for a specified price by a specified date. Options are primarily used to hedge against future risks or, perhaps more frequently, as a means of speculation.

The first reported use of options[9] seems to be by Thales who, after predicting a large olive crop by astronomical or astrological means, purchased one winter the right to use all the olive presses in Miletus and Chios for the coming harvest. When the large crop materialized, Thales was able to resell his rights for much more than they had cost him. Later on, tulip bulb options were heavily traded in Holland during the tulip mania of 1636 (until the market collapsed in 1637). Options were first traded on an organized exchange in 1973, on the Chicago Board Options Exchange. Since then the trading of options has developed into a global financial market of enormous volume.

[9]Aristotle, *Politics* I xi, 332 B.C.

There are two main types of options: a call option gives the holder the right to buy the underlying asset, while a put option give the holder the right to sell the asset. In an American option this right can be exercised at any time up to the expiration date; in a European option, the right can be exercised only on the expiration date itself.

Any options contract has two parties. The party who buys the option, is said to take the long position, while the party who sells, or writes, the option is said to take the short position. The writer receives cash up front, but has potential liabilities later on if the holder exercises the option. The holder incurs an immediate cost, but has the potential for future gains.

Let us consider, as an example, a European call option which gives the holder the right to buy a unit of stock at a prearranged price $K > 0$, called the *strike price*, at a future time $T > 0$. In this case, the value, or payoff, of the option at the expiration time T for stock price S is

$$f(S,T) = \max\{S - K, 0\}. \tag{5.75}$$

If $S \leq K$, the option is worthless, and the holder lets it expire; if $S > K$, the holder exercises the option and makes a profit equal to the difference between the actual price of the stock at time T and the strike price. We want to compute the fair value of the option at an earlier time.

To do this, we solve the Black-Scholes PDE (5.74) for $t \leq T$ subject to the final condition (5.75). We can find the solution explicitly by transforming (5.74) into the heat equation.

The change of independent variables $(S,t) \mapsto (x,\tau)$ given by

$$S = Ke^x, \qquad t = T - \frac{1}{\sigma^2}\tau \tag{5.76}$$

transforms (5.74) into the constant-coefficient equation

$$\frac{\partial f}{\partial \tau} = \frac{1}{2}\frac{\partial^2 f}{\partial x^2} + q\frac{\partial f}{\partial x} - \left(q + \frac{1}{2}\right) f \tag{5.77}$$

where

$$q = \frac{r}{\sigma^2} - \frac{1}{2}. \tag{5.78}$$

Since $0 < S < \infty$, we have $-\infty < x < \infty$. We have also reversed the time-direction, so that the final time $t = T$ corresponds to the initial time $\tau = 0$. The change of dependent variable in (5.77)

$$f(x,\tau) = Ke^{-qx-(q+1)^2\tau/2}u(x,\tau) \tag{5.79}$$

gives the heat equation

$$\frac{\partial u}{\partial \tau} = \frac{1}{2}\frac{\partial^2 u}{\partial x^2}. \tag{5.80}$$

Rewriting (5.75) in terms of the transformed variables, we get the initial condition

$$u(x,0) = \begin{cases} e^{(q+1)x} - e^{qx} & \text{if } x > 0, \\ 0 & \text{if } x \leq 0. \end{cases} \tag{5.81}$$

The Green's function representation of the solution of (5.80)–(5.81) is

$$u(x,\tau) = \frac{1}{\sqrt{2\pi\tau}} \int_0^\infty \exp\left[-\frac{(x-y)^2}{2\tau}\right] \left[e^{(q+1)y} - e^{qy}\right] dy.$$

This integral is straightforward to evaluate by completing the square. For example,

$$\begin{aligned}\int_0^\infty \exp\left[-\frac{(x-y)^2}{2\tau}\right] e^{qy}\,dy &= \exp\left[qx+\frac{1}{2}q^2\tau\right]\int_0^\infty \exp\left[\frac{(y-x-q\tau)^2}{2\tau}\right]\,dy\\ &= \sqrt{\tau}\exp\left[qx+\frac{1}{2}q^2\tau\right]\int_{-\infty}^{\left(\frac{x+q\tau}{\sqrt{\tau}}\right)} e^{-z^2/2}\,dz\\ &= \sqrt{2\pi\tau}\exp\left[qx+\frac{1}{2}q^2\tau\right]\Phi\left(\frac{x+q\tau}{\sqrt{\tau}}\right)\end{aligned}$$

where Φ is the distribution function of the standard Gaussian,

$$\Phi(x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-z^2/2}\,dz.$$

The function Φ is given in terms of the error function erf by

$$\Phi(x) = \frac{1}{2}\left[1+\operatorname{erf}\left(\frac{x}{\sqrt{2}}\right)\right], \qquad \operatorname{erf}(x) = \frac{2}{\sqrt{\pi}}\int_0^x e^{-z^2}\,dz.$$

It follows that

$$\begin{aligned}u(x,\tau) = \exp\left[(q+1)x+\frac{1}{2}(q+1)^2\tau\right]\Phi\left(\frac{x+(q+1)\tau}{\sqrt{\tau}}\right)\\ -\exp\left(qx+\frac{1}{2}q^2\tau\right)\Phi\left(\frac{x+q\tau}{\sqrt{\tau}}\right).\end{aligned}$$

Using this equation in (5.79), then using (5.76) and (5.78) to rewrite the result in terms of the original independent variables (S,t), we get

$$f(S,t) = S\Phi\left(a(S,t)\right) - Ke^{-r(T-t)}\Phi\left(b(S,t)\right) \tag{5.82}$$

where

$$\begin{aligned}a(S,t) &= \frac{1}{\sigma\sqrt{T-t}}\left[\log\left(\frac{S}{K}\right)+\left(r+\frac{1}{2}\sigma^2\right)(T-t)\right],\\ b(S,t) &= \frac{1}{\sigma\sqrt{T-t}}\left[\log\left(\frac{S}{K}\right)+\left(r-\frac{1}{2}\sigma^2\right)(T-t)\right].\end{aligned} \tag{5.83}$$

Equation (5.82)–(5.83) is the Black-Scholes formula for the value $f(S,t)$, at stock-price S and time t, of a European call option with strike price K and expiration time T. It also involves the risk-free rate of return r of the market and the volatility σ of the underlying stock.

Other types of options can be analyzed in a similar way. American options are, however, more difficult to analyze than European options since the time, if any, at which they are exercised is not known *a priori*.